The efficiency and accuracy of a rapid qualitative tool to ascribe socio-economic status in communities in South Africa – a cross-sectional study.

by Melissa Nel

Dissertation presented for the Degree of Master of Philosophy (Transdisciplinary Health and Development Studies) in the Faculty of Arts and Social Sciences, at Stellenbosch University

Supervisor: Dr Graeme Hoddinott

Co-supervisor: Dr Rory Dunbar

December 2020

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own original work, that I am the authorship owner thereof (unless to the extent explicitly otherwise stated) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: 27 November 2020

Abstract

Socio-economic status (SES) is a well-established construct. Lower SES is consistently associated with increased health challenges. SES is important to social policy and health interventions and, therefore, constant effort is made to improve its measurement.

I identified the varied practices of standard SES scale construction and measurement. The plethora of different scales and measures creates research inconsistencies. Validity and reliability are challenges for many SES scales, especially in lower to middle income countries (LMICs). Due to the lack of a generalised SES scale, cross-comparison in different contexts comes with many caveats. Additionally, difficulties are experienced when carrying out research that deals with the collection of large-scale data. The data collection process is labour intensive, time consuming, expensive, and complicated by differing norms and economic systems.

I explored an alternative SES measuring process that is quicker and more operationally useful for health intervention and policy planning. This approach is called the "Qualitative Ascription of SES (QASES)" in which data are collected rapidly and observationally and then SES is ascribed to local neighbourhoods by the research staff.

My data analysis was exploratory and comparative of secondary data that were collected using the QASES measure and a standard SES survey at individual-level. Firstly, I ran experiments to determine the efficiency of QASES compared to an individual-level SES survey. I created hypothetical contextual scenarios of a small study area and a large study area. I applied both methods to the study areas and determined the data collection processes in terms of labour, costs and time requirements. Secondly, I applied correlation analysis (Spearman's rho) to the existing data where QASES and a standard SES survey was used in 9 study communities of South Africa. I determined the strength of associations between the QASES scale and a standard SES survey. I estimated that QASES is approximately 1.5x cheaper and 2x faster to implement than an individual-level SES survey, which makes QASES more operationally useful. In addition, the correlation between QASES and the standard SES measure showed a strong, positive association ($r=0.753$, $n=142$, $p=0.000$). Therefore, I found that the QASES approach can be used as a substitute for standard SES data

collection, especially in LMICs. I recommend that the study should be replicated to further develop the QASES tool.

Opsomming

Sosio-ekonomiese status (SES) is ’n gevestigde konstruksie. Laer-SES word konsekwent met verhoogde gesondheidsuitdagings geassosieer. SES is belangrik vir sosiale beleid- en gesondheidsintervensies en daarom word voortdurend gepoog om die meting daarvan te verbeter.

Ek het die verskillende praktyke van SES-skaalkonstruksie en -meting geïdentifiseer. Die oorvloed van verskillende skale en maatstawwe skep teenstrydighede met die navorsing. Geldigheid en betroubaarheid is uitdagingend vir baie SES-skale, veral in laer- tot middelinkomste-lande (LMIL). Vanweë die gebrek aan ’n veralgemeende SES-skaal, kom kruisvergelyking in verskillende kontekste met baie voorwaardes voor. Boonop word probleme tydens die uitvoering van navorsing, wat handel oor die versameling van grootskaalse data, ondervind. Die data-insamelingsproses is arbeidsintensief, tydrowend, duur en ingewikkeld as verskillende norme en ekonomiese stelsels in ag geneem word.

Ek het ’n alternatiewe SES-meetproses ondersoek wat vinniger en meer bruikbaar is vir gesondheidsintervensie en beleidsbeplanning. Hierdie benadering word die “kwalitatiewe toeskrywing van SES (QASES)” genoem waarin data vinnig en waarnemend versamel word. SES word dan aan plaaslike woonbuurte deur die navorsingspan toegeskryf.

Ek het sekondêre data op ’n verkennende en vergelykende wyse ontleed. Hierdie data is met behulp van die QASES-maatstaf en ’n individuele standaard-SES-opname ingesamel. Eerstens het ek eksperimente uitgevoer om die doeltreffendheid van QASES in vergelyking met ’n individuele vlak SES-opname te bepaal. Ek het hipotetiese kontekstuele scenario's van ’n klein en groot studiegebied geskep. Albei metodes is op die studiegebiede toegepas en die proses van data-insameling is ten opsigte van arbeid, koste en tydsvereistes bepaal. Tweedens het ek korrelasie-analise (Spearman's rho) op die bestaande data toegepas waar QASES en ’n standaard-SES-opname in 9 studiegemeenskappe van Suid-Afrika gebruik is. Ek het die sterkte van assosiasies tussen die QASES-skaal en ’n standaard-SES-opname bepaal. Daar is beraam dat QASES ongeveer 1.5x goedkoper en 2x vinniger is as ’n individuele vlak SES-opname om te implementeer, wat QASES meer bruikbaar maak. Daarbenewens het die korrelasie tussen QASES en die standaard SES-maatstaf ’n sterk, positiewe assosiasie getoon ($r = 0.753$, $n = 142$, $p = 0.000$). Daarom kan QASES as ’n plaasvervanger vir standaard SES-data-

insameling gebruik word, veral in LMIL. Ek beveel aan dat die studie herhaal word om die QASES-instrument verder te ontwikkel.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Dr Graeme Hoddinott, for his patient guidance, continuous support and for never stop believing in me. I am grateful for my co-supervisor, Dr Rory Dunbar, who was always ready to offer me technical data support.

My appreciation goes out to everyone at the Desmond Tutu TB Centre who provided me with pivotal information, guidance and support throughout my studies. Also, thanks to my colleagues and friends at the centre for keeping me sane with their emotional and intellectual support.

I sincerely thank the Desmond Tutu TB Centre and the National Research Fund for their resources to conduct this research.

Finally, thank you to my family, friends and partner for help shaping me into the person I am today.

Table of Contents

Contents

List of Figures

List of Tables

List of Appendices

List of Abbreviations

AIDS	Acquired Immunodeficiency Syndrome
BBS	Broad Brush Survey
CRT	Cluster-randomised trial
DTTC	Desmond Tutu TB Centre is a research centre in the Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, Stellenbosch University
QASES	Qualitative Ascription of SES
HIV	Human Immunodeficiency Virus
HPTN 071 (PopART)	HIV Prevention Trial Network 071 (Population Effects of Antiretroviral Therapy to Reduce HIV Transmission)
PC	Population Cohort; the primary outcome of measuring HIV incidence in the HPTN 071 (PopART) trial
PC0	Population Cohort at baseline
PrEP	Pre-exposure prophylaxis
RDP	Reconstruction and Development Programme
SES	Socio-economic status
StatsSA	Statistics South Africa

Chapter 1: Introduction and Literature Review

1.1. Background

1.1.1. Embedded research

HPTN 071 (PopART) (Population Effects of Antiretroviral Therapy to Reduce HIV Transmission) was a cluster-randomised controlled trial in South Africa and Zambia conducted from 2013 through 2018 (Hayes, Donnell, Floyd, Mandla, Bwalya, Sabapathy, Yang, Phiri, Schaap, Eshleman, Piwowar-Manning, Kosloff, James, Skalland, Wilson, Emel, Macleod, Dunbar, Simwinga, Makola, Bond, Hoddinott, Moore, Griffith, Sista, Vermund, El-Sadr, Burns, Hargreaves, Hauck, Fraser, Shanaube, Bock, Beyers, Ayles & Fidler, 2019). In cluster-randomised trials (CRT) the focus is on comparing health interventions that are distributed randomly towards entire communities or intact clusters rather than individual research subjects (Hayes & Bennett, 1999; Lorenz, Köpke, Pfaff & Blettner, 2018). The trial was implemented to measure human immunodeficiency virus (HIV) incidence and prevalence in the two sub-Saharan African countries (Hayes, Ayles, Beyers, Sabapathy, Floyd, Shanaube, Bock, Griffith, Moore, Watson-Jones, Fraser, Vermund, Fidler, Agyei, Baldwin, Barnes, Bond, Burns, Chishinga, Cummings, Donnell, Emel, Eshleman, Godfrey-Faussett, Greene, Hargreaves, Hauck, Headen, Horn, Kim, Piwowar-Manning, McCarthy, Musheke, Mwango, Mwinga, Muyoyeta, Simwinga, Schaap, Smith, Wolf & White, 2014). HIV incidence refers to the estimated number of new infections per annum, compared to HIV prevalence that indicates the estimated percentage of the total population living with HIV (UNAIDS, 2015).

In adults aged 15 to 49, the national HIV prevalence at the time of planning the trial was 13.5% in Zambia and 17.8% in South Africa, whereas estimates for incidence were 1.06% and 1.49%, respectively (Hayes, Ayles, Beyers & Sabapathy, 2014). The trial evaluated the effect of a combination HIV prevention package on HIV incidence. Part of the prevention package was making universal testing and treatment (UTT) available in intervention study communities. By intervention, the notion is that a few communities out of the total are selected to receive UTT as a HIV prevention package with the purpose of curbing HIV prevalence. UTT is aimed at maximising HIV testing in

populations, combined with effective linkage to care and the immediate onset of antiretroviral treatment (ART) (Hayes, Ayles, Beyers, Sabapathy, *et al.*, 2014).

In preparation of the commencement of the trial, formative research was conducted in the study communities through an approach called "Broad Brush Surveys" (BBS) in 2013. Social scientists rapidly observed and evaluated the communities to understand how the community is structured, to identify important stakeholders, and to gain information on the impact of HIV (Bond, Hoddinott, Musheke, Viljoen, Abrahams, Chiti, Mantantana, Ndubani, Simuyaba & Seeley, 2013). Infrastructure, facilities and sub-neighbourhoods were observed and interviews or group discussions were held with community members. Data was collected that captured the study communities' socio-economic status (SES), whereby researchers qualitatively observed specific features and quantitatively ascribed SES scores to these features. To evaluate the effect of the combination prevention intervention on HIV incidence and viral suppression, a population cohort was implemented during the trial to enrol and follow a representative sample of residents from 2014 to 2018 (Hayes *et al.*, 2019). These individuals were enrolled from randomly selected households and the first visit entailed the completion of a baseline survey, which included a section of questions to measure individual-level SES.

In this study, I illustrate and compare two examples of how SES data was collected during HPTN 071 (PopART): the novel SES data collection approach and an individual-level SES measure used in the population cohort. This secondary analysis of the data is done to identify whether the qualitative SES approach can yield similar results than a standardised measure of SES. Through a literature search, I draw on the integration of socio-economic status and health by providing context to different spheres that intersect and influence SES measurement. Specifically, on how race, gender, locality and poverty shape the occurrence of HIV as an example of a multifaceted illness/disease. Further, I discuss in detail what is known about existing SES measurement tools to emphasise the complexity and different ways of measuring SES in different contexts.

1.2. Socio-economic status and health

Socio-economic status (SES) is well-established as a determinant of health and has been emphasised extensively in research (Aggarwal, Bhasin, Sharma, Chhabra, Aggarwal & Rajoura, 2005; Cutler, Lleras-Muney & Vogl, 2008; Oakes & Rossi, 2003). Since SES attempts to capture complex information of one's life, studies continue to link this information to disease or disability (Oakes & Rossi, 2003). Definitions of SES vary, but according to Shavers (2007) the fundamental constant is having access to 'basic' resources. SES in relation to health is further refined as "influencing the accessibility, affordability, acceptability and actual utilization of various available health facilities" (Aggarwal *et al.*, 2005:111). The greatest health challenges are experienced by those who tend to experience the greatest socio-economic disparities, like ethnic minorities, the elderly and the young (Shavers, 2007).

It has long been argued by researchers like Kaplan, Haan, Syme, Minkler and Winkelby (1987) that globally, people at the lowest levels of SES have higher illness and death rates, regardless of what the major causes of disease or death are and how SES is measured. SES is associated with an extensive array of non-communicable health issues, including heart disease, stroke, cancer, diabetes, hypertension, infant mortality, injuries, poor nutrition, mental illness, and communicable diseases like HIV, TB, chicken pox, pneumonia, and diarrhoea (Anderson & Armstead, 1995; Ncho & Wright, 2013). There exists a continuous focus on SES to help predict future prognosis of disease and explain how co-morbidities and co-infections occur (Anderson & Armstead, 1995; Ferreira Antunes, Waldman & Borrell, 2005; Glanville, et al., 2019).

1.2.1. Race, SES and health

In my dissertation I use the following race labels: black, white, coloured, and Indian. These racial categories are social constructs and not essential truths, but race remains an important predictor of both health and SES. I follow the trajectory of Finchilescu & Tredoux (2010:228) that "these groups have a historical reality that has shaped the subjectivities and worldviews of the South African population". In South Africa, "disparities in wealth and health are among the widest in the world" (Benatar, 2013:1). These inequities are rooted in the country's history of policies deriving from the

periods of colonialism, apartheid and post-apartheid (Coovadia et al., 2009). Despite some changes, the country's contemporary infrastructural landscape still reflects spatial engineering of resources along a racial hierarchy benefiting the white minority (Coovadia, Jewkes, Barron, Sanders & McIntyre, 2009; Seekings, 2010). Even though the emergence of a democratic society has led to the abolishment of all discriminatory laws and practices, black, coloured, and Indian people have not experienced radical changes in terms of material well-being (Finchilescu & Tredoux, 2010).

In South Africa, race is therefore intertwined with both locality and SES. Out of approximately 1.9 million people living in informal dwellings, often in urban/peri-urban informal 'settlements', around 1.6 million of them are black (Statistics South Africa, 2016). According to Lombard (2014), urban informal settlements are typically defined by certain criteria, such as low incomes of residents, self-build housing made from scrap materials and sub-standard infrastructure and services. Furthermore, the total annual consumption expenditure (how much is spent on goods and services) was 3.7% in urban informal settlements compared to 82.2% for urban formal settlements (Statistics South Africa, 2015). The health of the majority of South Africans are negatively influenced by persistent non-communicable and infectious diseases, ongoing social inequalities and a deficiency of human resources to deliver care (Mayosi & Benatar, 2014). For instance, Ataguba et al. (2011:4) found that the bottom 40% (poor quintiles) of South Africa's population, "bears about 56% of the burden [of HIV] compared to 11% for the top 40% (rich quintiles)".

1.2.2. HIV epidemiology and SES in South Africa

One of the biggest health issues and causes of mortality in South Africa is the HIV epidemic. Of the total world population affected by HIV, approximately 20% of them live in South Africa (Gutreuter, Igumbor, Wabiri, Desai & Durand, 2019; Kevany, Benatar & Fleischer, 2013; UNAIDS, 2018). Prevalence is highly unequal by race, sex, age, locality type and province and these factors combined are intertwined with SES (Bunyasi & Coetzee, 2017; Shisana, Rehle, Simbayi, Zuma, Jooste, Zungu, Labadarios, Onoya & Al., 2014). Most black South Africans face extreme hardships with "high levels of unemployment, lack of housing, inadequate education, poor levels of health care, and the scourge of HIV/AIDS" (Finchilescu & Tredoux, 2010:226). For instance, HIV is a disease

embedded in socio-economic inequity which disproportionately affects those living in lower socio-economic communities (Bunyasi & Coetzee, 2017). Prevalence of HIV by locality is 19.9% for urban informal settlements compared to 10.1% for urban formal settlements (Shisana *et al.*, 2014). HIV is therefore a good example of the close and complex relationship between SES and health.

1.2.3. SES, HIV and poverty

During the early years of the HIV epidemic in sub-Saharan Africa, HIV was more prevalent among the relatively wealthy due to their abundance of disposable income and engagement in multiple sexual partnerships (Wabiri & Taffa, 2013). HIV prevalence was also more likely to be diagnosed among the wealthy due to their higher accessibility to healthcare and hence testing for HIV (Smart, 2006). As the epidemic matured, those in the poorer income brackets became equally affected as sexual networks expanded (Wabiri & Taffa, 2013). HIV affected the poor more severely due to "lost economic opportunities and cost of caring" (Wabiri & Taffa, 2013:1). In South Africa, these patterns were skewed by institutionalised and then normative restrictions on inter-racial mixing and sex (Shisana, Zungu & Pezi, 2009). Therefore, the context specific socio-economic impact of HIV became strongly associated with poverty when HIV expanded. Besides for poverty, HIV incidence in sub-Saharan Africa were most strongly related to socio-economic inequality and vulnerability (Wabiri & Taffa, 2013). Having less education and wealth or experiencing greater levels of poverty are associated with higher transmission rates of HIV (Bunyasi & Coetzee, 2017).

Many people in South Africa experience poverty based on "deprivation, constrained choices, and unfulfilled capabilities" that directly impacts their quality of life and standard of living (Mbirimtengerenji, 2007:605). Together with a deficiency of money, there's a lack of skills and assets (Mbirimtengerenji, 2007). For instance, school dropouts have a higher risk of obtaining HIV as they engage more commonly in intergenerational sex, have a higher number of sexual partners, and engage more often in unsafe sex, compared to those who complete school (Bunyasi & Coetzee, 2017). HIV in turn contributes to the severity of poverty by burdening the household's expenditure on medical costs, or a family member that becomes unable to provide for the household, therefore a loss of income (Bunyasi & Coetzee, 2017). SES helps to understand the capacity of individuals and households to

cope with HIV when looking at their endowment of assets and resources – both human and financial (Mbirimtengerenji, 2007).

1.3. Conceptualising and measuring SES

1.3.1. Defining SES and its indicator components

SES is understood "as the social standing or class of an individual or group, often measured as a combination of education, income, and occupation" (Berzofsky, Smiley & Krebs, 2014:2). The latter three indicators are considered the traditional variables in measuring SES and are often standardised at individual, family or household level (Berzofsky *et al.*, 2014). Developing a valid and reliable SES measurement scale is a research priority (Tiwari, Kumar & Kumar, 2005). Current scales have been critiqued as outdated (Oakes & Rossi, 2003) and some variables need to be redefined to indicate SES more accurately (Milenkovic, Vukmirovic, Bulajic & Radojicic, 2014). Data representation are flawed where little agreement exists on which SES indicators should be grouped and collected, despite a growing awareness of the need to collect SES indicators regularly (Duncan, Daly, McDonough & Williams, 2002).

The mentioned indicators are not interchangeable, meaning that varied socio-economic factors can influence health differently during the life course, working at different levels (e.g., individual or neighbourhood) and through different contributory pathways (e.g., environmental exposures or vulnerability) (Braveman, Cubbin, Egerter, Chideya, Marchi, Metzler & Posner, 2005; Pollack, Chideya, Cubbin, Williams, Dekker & Braveman, 2007). SES indicators are influenced by covariate factors like sex or age that varies across different population sub-groups. For instance, the use of occupation as an indicator in studies involving women has been problematic. Standard occupational systems tend to discriminate along the lines of gender-based occupations (Shavers, 2007). A variety of SES indicators capture different facets of health risk which becomes challenging when deriving optimal indicators to measure SES (Duncan *et al*., 2002). Specifically, in lower-/middle-income countries (LMICs) where the standard measurement of SES is less representative to the diverse circumstances that each country faces. For instance, in low-income communities it is challenging to

collect income data due to informal work and monthly fluctuations in work (Psaki et al., 2014). This has resulted in replacing income information with that of measuring accumulated wealth by household assets as an index of household materials[1] (Psaki, Seidman, Miller, Gottlieb, Bhutta, Ahmed, Ahmed, Bessong, John, Kang, Kosek, Lima, Shrestha, Svensen & Checkley, 2014). Assets would indicate more informative trends than income and produces a representative meaning for the groups under study (Shavers, 2007).

1.3.2. Categories of SES measurements

Measuring SES can take place at different levels and have different indicators, depending on the available data and the study design employed (Braveman, et al., 2005; Berzofsky, et al., 2014). The three complementary levels of SES measurement are: individual, household and neighbourhood (Krieger, Williams & Moss, 1997). Each level can contribute to outcomes or exposure distributions independently (Krieger et al., 1997). Moreover, all the relevant categories of SES measurement are identified from a literature search (see Figure 1). According to Shavers (2007) and Oakes (2008), the two basic approaches to SES and health are compositional and contextual. The former is applicable to the individuals' socio-economic and behavioural characteristics. The latter examine the socio-economic conditions of the environment shared by individuals (Shavers, 2007). Typically, compositional and contextual indicators are not measured separately but either as a composite or at multi-level (Shavers, 2007). By composite, the information of several SES measures/indicators are combined (e.g., income, occupation, housing, employment, area-level) and are measured either at individual, household or family level (Shavers, 2007).

Multi-level analysis places a lot of emphasis on the context in combination with compositional measures (Shavers, 2007). The distinction is that a multi-level approach is used to measure different SES indicators at various levels, for instance, individual, family and neighbourhood levels (Yang & Gustafsson, 2010). Whereas a composite measure might use combined compositional indicators (e.g., education and income) applied at individual area-level. The use of a single individual measure of SES

[1]Asset index are used as a proxy for substituting income or expenditures variables, and captures household belongings (electricity, oven, stove, radio, refrigerator, TV, bicycle, motorcycle, car, and telephone). Household materials include source of drinking water, toilet facilities and flooring material (Fotso & Kuate-Defo, 2005).

may result in absent individual information. Neighbourhood or contextual variables can act as proxies to fill the gaps (Pickett & Pearl, 2000).

Furthermore, each type of SES can be measured subjectively or objectively. Objective SES is the socio-economic position of an individual, family or household relative to others (Demakakos et al., 2008). In contrast, subjective SES is an individual's experience of their position compared to other individuals (Huang et al., 2017). In most instances there is a reliance on objective SES measures (e.g., education, occupational class, and wealth/income) and these measures are standardised to account for consistency and reliability (NCVHS, 2012). However, subjective SES can be used as a "potential mediator of the associations between objective indicators of SES and health" (Demakakos, et al., 2008:331). People get a chance to assess their own deprivation experiences and social status perceptions through subjective SES (Singh-Manoux, et al., 2003; Demakakos, et al., 2008).

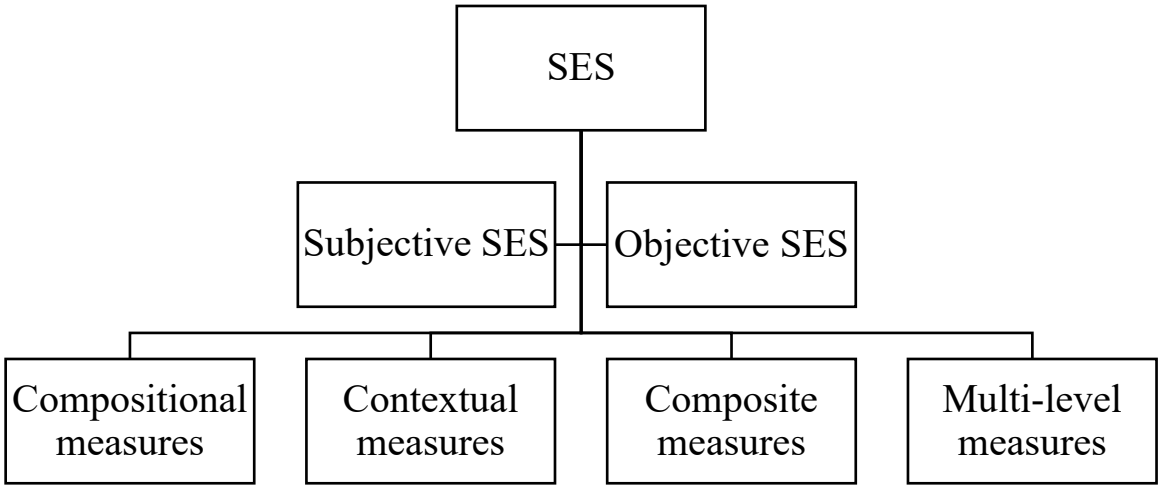


Figure 1: SES and its different categories of measurement

1.3.3. Indicators of SES

There is no particular indicator of SES that is most appropriate for application across settings (Galobardes, Lynch & Smith, 2007). Rather, each indicator measures different but related aspects of SES (Fliesser, De Witt Huberts & Wippert, 2018). A chosen or preferred indicator should be justified as each indicator has a different influence on the association with a specific health outcome (Fliesser

et al., 2018). Typically, individual-level indicators of SES measure some type of resource or asset of an individual (Galobardes *et al.*, 2007).

Of the three traditional SES indicators, income and/or wealth measure material circumstances by the gross/annual household income, family income or individual annual income (Galobardes *et al.*, 2007; Shavers, 2007). Most often the government/official poverty level are used as a reference point to define income categories (low, medium and high) by either dividing them into tertiles or quintiles (Berzofsky *et al.*, 2014). Income provides an idea of how the extent of access to material goods, like food and shelter, and access to services at a particular time can influence health (Galobardes *et al.*, 2007). Wealth captures information on the accumulated resources and can be regarded as an extension to income (Galobardes *et al.*, 2007; Shavers, 2007). There is a distinction between how wealth is used as an indicator of SES depending on the context, specifically across high-income countries (HICs) and LMICs. In HICs measures of wealth can include assets (estimated cash value of an individual's home, property or similar investments) or net worth (total assets minus outstanding debt) (Pollack *et al.*, 2007). In LMICs wealth is measured by the use of a wealth index (also referred to as an asset index or standard of living index) (Howe, Hargreaves, Ploubidis, De Stavola & Huttly, 2011). Typically, the wealth index is a composite measure that encompass "ownership of consumer durables, access to services and dwelling characteristics" (Howe *et al.*, 2011: 224).

Education is regarded as an extensive indicator of SES because it influences earning potential and occupational opportunities across the lifespan (Berzofsky *et al.*, 2014; Shavers, 2007). It is also greatly associated with knowledge around health and available treatments (Fliesser *et al.*, 2018). Lower risk of HIV infection has been linked to an increased educational achievement due to an improved "ability to understand and act on health promotion messages, [an] increased exposure to school-based HIV prevention programmes or increased access to health services" (Bärnighausen, Hosegood, Timaeus & Newell, 2007: 4). Education is measured by looking at "years of education completed, highest educational level completed, and credentials earned (e.g., High school certificate, Bachelor's Degree)" (Shavers, 2007:1015).

Occupation as indicator of SES, entails a person's income, power and educational requirements to acquire positions in the occupational structure (Berzofsky *et al.*, 2014). The influence of occupation on health are determined by the occupational rank or social class, and circumstances of the physical work environment (Shavers, 2007). Therefore, occupation can act as an intermediate indicator that can affect and be affected by both income and education (Omer & Al-Hadithi, 2017). Occupation categories are specified by ranked labour from lowest to highest SES (e.g., Unemployed, Unskilled Manual Labour, Skilled Manual Labour, and Professional Labour). Another approach is to rank occupations on a scale from 1 to 100 for their perceived prestige which is termed occupational prestige scaling. Rankings are confirmed from surveys where respondents are asked to rank their own or other occupations (LaVeist, 2005).

Area-level indicators of SES "evaluates the geographical distribution of socio-economic inequalities in health" (Galobardes *et al.*, 2007: 31). Area or contextual SES measures, like the one explored in this thesis, are designed to represent the individual's environment that look at neighbourhoods (identified via census tracts, census blocks and postal codes) to larger areas like districts and provinces (Berzofsky *et al.*, 2014). Commonly used area-based measures of SES include average home value, amount of higher educated people, percentage of unemployment, and single-parent households. These can be used as single items or combined into scales. The value of contextual-level SES is that it provides a substitution for income data that are typically absent – due to a high non-response rate – in survey data at individual or household level.

In addition to the traditional measures, many proxy indicators exist that offer a valuable approach to SES. Proxy indicators are used when direct or traditional measures are not available (Galobardes, Shaw, Lawlor, Lynch & Smith, 2006a). Proxy indicators that might yield valuable to SES measurement are housing (housing tenancy, household facilities, household characteristics), unemployment, overcrowding (calculated by the number of people living in a household per bedroom), infant and maternal mortality (ecological measures of an area or country), and area-level indicators (Galobardes *et al.*, 2007, 2006a). Housing proxies are usually combined with an asset index to assess the relative wealth of a household within a population (Howe, Hargreaves & Huttly, 2008).

For instance, household facilities and amenities such as water accessibility, electricity, whether the toilet is situated outside or inside the house, and type of walls and flooring are combined with household assets such as the ownership of a TV, washing machine, stove, and refrigerator (Galobardes *et al.*, 2006a; Howe *et al.*, 2008).

Combining household amenities and assets into one proxy measure can also be described as a wealth index and are particularly popular to use in LMICs as a substitute to income and household consumption and expenditure (Howe, Galobardes, Matijasevich, Gordon, Johnston, Onwujekwe, Patel, Webb, Lawlor & Hargreaves, 2012). The majority of wealth indices that are employed to an array of topics are created using a method where principal components analysis (PCA) is used to summarise multi-dimensional information on various household assets ownership (Filmer & Scott, 2012; Poirier, Grépin & Grignon, 2020). Using Demographic Health Surveys (DHS) to measure household wealth by applying PCA "allowed researchers to convert a series of ownership variables, many of which are binary (yes/no) or categorical (roof material, housing types, etc.) into a continuous SES gradient" (Poirier *et al.*, 2020: 2). The wealth index construction using PCA has become the standardised proxy for household SES as a substitute of income and consumption data (Poirier *et al.*, 2020).

1.3.4. Notable SES indices in HICs and LMICs

In Table 1, Table 2 and Table 3 all relevant scales are listed with a description of each. The scales are grouped as compositional, contextual and composite measures. It is worth mentioning that many of the scales listed are considered as original of which multiple other scales have been developed from. Many of the scales are also considered to be outdated or have been adapted to fit contemporary circumstances with regards to how income, education and occupation are perceived (e.g., prior to 1990 the registrar general's social class, now known as the British occupational-based social class) (Galobardes *et al.*, 2006a).

It is noted that measuring SES is different depending on the context, especially between HICs and LMICs (Bärnighausen *et al.*, 2007; Howe *et al.*, 2008; Ichoku, 2011). Therefore, a distinction should be made between scales that are typically used in HICs and LMICs. In HICs several SES

indexes have been created that are specifically used in health research. For instance, a composite SES measure such as Duncan's index (1961), categorises occupation according to income and education, where occupation is seen as an intervening variable between education as prerequisite and income as the reward (Oakes & Rossi, 2003). Occupational indices are also significantly (but not exclusively) used in European contexts where various types exist, especially in the United Kingdom (see table 1 for occupation indices) (Ichoku, 2011; Krieger, Williams & Moss, 1997). Townsend's index (1987) is a contextual measure designed to clarify how material deprivation influences area-level variation in health indicators (Morris & Carstairs, 1991). In the United Kingdom, Townsend's index is one of the most widely used measures for deprivation and identifies an area's population rate of unemployment, non-car and non-homeowners, and living in overcrowded households (Krieger *et al.*, 1997). In LMICs, measures are often more dependent on proxy indicators as alternative measures of SES based on income, consumption and educational attainment (Ichoku, 2011). The few attempts made to create a SES asset index based on housing quality indicators such as:

> wall and roofing material, cooking and lighting fuel, source of drinking water, sewage system, and tenure, on household wealth, housing, education and occupation, or on a broader sequence of familial living conditions namely housing, literacy and cultural aspects, demographic conditions, economic conditions (Fotso & Kuate-Defo, 2005: 192).

Household assets indices as proxies have shown to be valid for wealth measurement in rural Africa and are increasingly being used in primary data collection in LMICs (Bärnighausen *et al.*, 2007; Howe *et al.*, 2012).

The United Nations Development Program (UNDP) developed the Human Development Index, which captures the average of the SES measurement in three dimensions: "longevity indicator based on life expectancy at birth, educational attainment based on the percentage of the literacy of the adult population and the children's school enrolment, and resource indicator based on the per capita gross domestic product (GDP)" (Fotso & Kuate-Defo, 2005: 192). These indexes are critiqued as rarely being developed to perform comparisons across LMICs even though it is relevant to establish and improve the linkages between SES and health in these contexts (Fotso & Kuate-Defo, 2005).

Bärnighausen *et al*. (2007) have developed a household assets index in South Africa based on house/property ownership, types of water sources, electricity, toilet type, energy and 27 household assets that may be used for both consumption and production (e.g., bicycles, telephones, beds, tables, televisions, sewing machines, block makers, tractors, wheelbarrows, cattle, and other livestock). Determined from the assets index scale, three categories of relative wealth are used to categorise households to belong to either the bottom 40%, the middle 40% or the upper 20% (Bärnighausen *et al.*, 2007). It has been found that wealth effects are captured well by these three categories in poor provinces of South Africa. Similarly, Statistics South Africa (StatsSA) developed a standard wealth index that gives a score based on the consumer goods owned by a household (e.g., car, television, number and type of livestock) and housing characteristics (e.g., flooring material, toilet facilities and drinking water source) (National Department of Health, 2017). A household score is assigned to each household member in order to rank the household population by their individual scores. This is to compile national wealth quintiles (i.e., five equal categories, each comprising 20% of the population). The lowest wealth quintile (bottom 20%) consists of the population with the fewest assets of least value, while the highest quintile (upper 20%) consists of the population with the most assets of greatest value (National Department of Health, 2017: 8).

1.3.5. Difficulties with large-scale SES data collection in resource-scarce settings

Using the standard approach to SES data collection, which is surveys, comes with many complications related to the required time, resources and management of data collection efforts. Across different contexts, these complications also vary in extent. SES data has been collected through paper-based questionnaires up until the turn of the 2010s (Seebregts, Zwarenstein, Mathews, Fairall, Flisher, Seebregts, Mukoma & Klepp, 2009; Walther, Hossin, Townend, Abernethy, Parker & Jeffries, 2011). Since then, paper-based surveys have mainly been phased out in preference of electronic data capture methods through surveys on tablets, laptops or other electronic devices (Seebregts *et al.*, 2009). Data collection through electronic devices have increased the ability to save time during data capture and analysis and led to more effective data quality through online data validation checks (Seebregts *et al.*, 2009). In terms of field data collection, electronic devices have

also been deemed affordable and robust as a method to collect survey data. This is especially true in LMICs which typically lack sufficient resources (Walther *et al.*, 2011).

In LMICs, multi-topic survey instruments are usually administered that encompass a section on SES like an asset or wealth index or other measures of poverty and SES, i.e., personal income, household income or GDP (Gross Domestic Product) per capita (Sweeney, Vassall, Foster, Simms, Ilboudo, Kimaro, Mudzengi & Guinness, 2016). This means that these surveys include different comprehensive sections causing them to be lengthy (Sweeney *et al.*, 2016). In health intervention research, survey length is of particular concern when a participant is asked to complete a questionnaire after a clinical investigation, as the risk for survey fatigue can be increased (Sweeney *et al.*, 2016). Either the interviewer and respondent could suffer from fatigue during lengthy surveys where surveys could be rushed to be completed, participants may refuse to continue, or resources should be increased to ensure the survey is conducted (Beegle, De Weerdt, Friedman & Gibson, 2012; Sweeney *et al.*, 2016). Interviewers or fieldworkers are also expected to complete multiple surveys per day, depending on the study sample size and further consideration of clustering and non-response (Sweeney *et al.*, 2016). This impacts the workload for data collectors who then must return to households or participants in order to meet the sampling requirements. Even though electronic data capture methods have reduced data collection errors, increased the reliability of data capturing, and facilitated in survey completion, it still requires training of fieldworkers on the devices (Walther *et al.*, 2011). Additional resources are required to train interviewers/fieldworkers on "data entry and security, and planning for power and connectivity issues" (Sweeney *et al.*, 2016: 48). These issues contribute to the complexities of SES data collection on top of the differing conceptualisations of SES and its measurement.

1.3.6. Qualitative approaches to SES measurement

From the literature search, I found no direct measures of SES using qualitative measures or approaches. Instead, qualitative research is usually done as a by-product of quantitatively determined SES studies using standard surveys. For instance, many studies focus on evaluating populations' health behaviours or educational aspirations based on pre-determined SES of populations, contexts or neighbourhoods (Berger & Archer, 2018; Eyre, Duncan, Birch & Cox, 2014; Roshita, Schubert &

Whittaker, 2012; Van Wijk, Overberg, Kunst & Harting, 2020). This is typically done through participatory methods like observational field notes, focus group discussions, and qualitative interviews with a number of participants from low SES and high SES backgrounds (as comparators) (Eyre *et al.*, 2014; Grant, Edwards, Sveistrup, Andrew & Egan, 2010).

Another common qualitative approach is to focus on health care providers' perspectives on patient care across different levels of SES (Bernheim, Ross, Krumholz & Bradley, 2008; Diniz, Castro, Bousfield & Figueira Bernardes, 2020). Bernheim *et al*., (2008) conducted semi-structured interviews with physicians in Connecticut (US) to elicit their caring practices for people of low SES. Diniz *et al*., (2020) made nurses, from several public and private hospitals in Lisbon and Porto (Portugal), watch short videos of two white women with similar pain levels from different SES (low and middle) and asked them to write a story of the women's pain, lives and treatment recommendations. Qualitative measures of SES are used as a means to strengthen standard SES data by providing context through narratives and observations. Instead of focusing on a numeric number that generalise an area or ethnic group's SES, qualitative measures provide descriptive accounts (Berger & Archer, 2018).

Some qualitative studies use quantitative SES measures to inform their sampling frame. For example, studies done by Roshita *et al*., (2012), Eyre *et al*., (2014) and Van Wijk *et al*., (2020) focused on interviewing the parents or caregivers of children to understand health behaviour practices such as child-care and feeding (Depok, Indonesia), children's physical activity (Coventry, UK) and second-hand smoke exposure (provincial town in the Netherlands) influenced by cultural, social and built environments. All three studies used residential locations as a proxy indicator of SES to select study areas for recruitment of participants to interview individually or in focus groups (Eyre *et al.*, 2014; Roshita *et al.*, 2012; Van Wijk *et al.*, 2020).

None of these uses are as described in my analysis where qualitative data are used to then ascribe quantitative scores on a scale. This makes my analysis highly novel, but also exploratory, with further empirical evidence required to further validate the approach.

Table 1: Indices and scales based on compositional measurements

Compositional scales ***Single measures at individual, household or family level***	
Indices	**Scales**
Occupational	1. **Registrar general's social class or British occupational-based social class** – (UK) "Groupings of occupation based on prestige in six hierarchical groups: I (highest), II, III non-manual, III-manual, IV, V (lowest). Often regrouped as manual versus non-manual" (Galobardes *et al.*, 2007: 27). 2. **Erikson and Goldthorpe class scheme** – (UK; industrialised societies) "Groupings of occupations based on specific characteristics of employment relations such as type of contractual agreement, independence of work, authority delegation, etc. Not a hierarchical classification" 27 (Galobardes *et al.*, 2007: 27). 3. **Wright's Social Class Scheme** – (UK; industrialised societies) Based on Marxist principle of relation to the means of production where people are categorised in terms of "three forms of exploitation: (a) ownership of capital assets, (b) control of organisational assets, and (c) possession of skills or credential assets" (Galobardes *et al.*, 2007: 27). 4. **Lombardi *et al* social class classification** – (Brazil) Based on Marx's theories and similar to Wright's classification. Six classified groups of occupations: "Under proletariat (unemployed and seasonal workers); Typical proletariat (unskilled and semiskilled workers in manual occupations); Atypical proletariat (unskilled and semiskilled in commerce and services);

	Traditional small bourgeoisie (self-employed, small business owners); New small bourgeoisie (university trained professionals); Bourgeoisie (large business owners)" (Galobardes *et al.*, 2006a: 96). 5. **Cambridge Social Interaction and Stratification Scale** – (UK; Universal) "Based on patterns of social interaction in relation to occupational groups" (Galobardes *et al.*, 2006a: 95). 6. **Standard International Occupational Prestige Scale (SIOPS)** – (Universal) "devised by taking survey information on prestige ratings given by respondents to samples of jobs and calculating averages within and across societies" (Connelly et al. 2016: 7) 7. **International Socio-Economic Index (ISEI) –** (Universal) "calculates scores for occupations based on their average profiles in terms of the income and educational qualifications held by their incumbents (with some adjustments for age profiles)" (Connelly et al. 2016:7) 8. **Occupational-based census classification** – (Universal) several country-specific socio-economic classifications (e.g., Edwards US census classification) (Galobardes *et al.*, 2006a). 9. **Siegel Prestige Scale (1971)** – (US) "based on the merger of three national surveys that obtained prestige ratings of 412 occupations" (Galobardes *et al.*, 2006a: 97).
Educational	1. **International Standard Classification of Education (ISCED)** – (Universal) "combines school and vocational education, scored from 0 (less than primary education) to 5 (tertiary education)" (Fliesser *et al.*, 2018: 2–3)

	2. **The National Assessment of Educational Progress** (**NAEP**) – (US) SES data collected through self-report from students (Grades 4, 8 and 12); including parental education attainment (for grade 8 and 12) (Cowan, Hauser, Kominski, Levin, Lucas, Morgan, Spencer & Chapman, 2013).
Income	1. **B G Prasad classification (1961) –** (India) a scale based on per capita monthly income (modified in 1968 and 1970) (Shaikh & Pathak, 2017: 998).

Table 2: Indices and scales based on area-level/contextual measures

Contextual scales ***Neighbourhoods: ZIP codes, census tracts, census block groups and census blocks*** ***Other geographic areas: examples include counties, districts and provinces***	
Indices	**Scales**
Housing	1. **"Broken window" index** – (US) measure "housing quality, abandoned cars, graffiti, trash, and public-school deterioration at the census block level in the USA" (Galobardes, Shaw, Lawlor, Lynch & Smith, 2006b: 9). 2. **"Social standing of the habitat"** – (US) "combined characteristics of the building, their immediate surroundings and the local neighbourhood of residential buildings" (Galobardes *et al.*, 2006b: 9).
Deprivation	1. **Townsend Deprivation Index** – (UK) four standardised variables: "the proportion of unemployed, households with no car, households that are not owner occupied and of households with overcrowding (more than one person per room)" 98 (Galobardes *et al.*, 2006a: 98). 2. **Carstairs deprivation index** – (UK) similar to Townsend, "unemployment rate among men aged 16 and over who are economically active, the percentage of non-car ownership among all households, household overcrowding and an economically active head of household in a deprived situation" (Galobardes *et al.*, 2006a: 98). 3. **Jarman or Underprivileged Area score** – (UK) similar to Townsend, an index to identify 'underprivileged' areas (Galobardes *et al.*, 2006a).

	4. **The Breadline Britain Index** – (UK) "combining survey with census data and using weights to account for the different probability that subgroups in the population will experience a particular type of deprivation – based on the proportions of: unemployed, people with no car, non-owner occupied households, lone-parent households, households with persons with long-term illness and unskilled and semi-skilled manual occupations (social class IV and V) in an area" (Galobardes *et al.*, 2006a: 98).

Table 3: Indices and scales based on composite measures

Composite scales ***At individual (usually measured as a score that adds up the presence or absence of several SES indicators) or at area level***	
Indices	**Scales**
Material and social deprivation	1. **Wealth Index (WI)** – "construction materials of dwelling houses and household assets are combined very common in LMICs" (Howe *et al.*, 2012: 872). 2. **Townsend Index** (see Table 2 for description) 3. **Carstrais Index** (see Table 2 for description) 4. **The Breadline Britain Index** (see Table 2 for description) 5. **Index of Multiple Deprivation** – (UK) "combines six domains: income, employment, health and disability, educational skills and training, housing and geographical access to services and was designed to measure various aspects of deprivation at ward level" (smallest unit in local governance) (Galobardes *et al.*, 2007: 32). 6. **Standard of Living Index (SLI) scale** – (India) contains 11 items: "housing type, source of lighting, toilet facility, main fuel for cooking, source of drinking water, separate room for cooking, ownership of the house, ownership of agricultural land, ownership of irrigated land, ownership of livestock, ownership of durable goods for measuring the SES both urban and rural areas for the entire country" (Kulkarni, Ramesh Masthi & Gangaboraiah, 2013: 69).

	7. **Bhuiya et al. SES scale –** (rural Bangladesh) "social involvement, food, clothing, education, shelter, and health as composite SES" (Saif-Ur-Rahman, Anwar, Hasan, Hossain, Shafique, Haseen, Khalequzzaman, Rahman & Islam, 2018: 2). 8. **Tiwari et al scale** – (India) seven indicators: "housing, material possession, education, occupation, monthly income, land, social participation and understanding" (Tiwari *et al.*, 2005: 309). 9. **Multidimensional Poverty Index (MPI)** – (South Asia) measured in over 100 countries and include ten indicators of "health (nutrition, child mortality), education (years of schooling, school attendance) and standard of living (cooking fuel, sanitation, drinking water, electricity, housing, assets)" (Saif-Ur-Rahman *et al.*, 2018: 2). 10. **Unsatisfied Basic Needs (UBN)** – (Latin America) i.e., access to clean water, housing quality, crowding, head of the household's level of education, school attendance, nutrition (Saif-Ur-Rahman *et al.*, 2018: 2). 11. **Human Development Index –** (Universal) **"**life expectancy, education, and per capita income indicators" (Milenkovic *et al.*, 2014: 604).
Social standing/prestige	1. **Hollingshead index of social position** – (US) four factors: "marital status, retired/employed status, educational attainment, and occupational prestige" (Galobardes *et al.*, 2006a: 98). 2. **Duncan's Socioeconomic index** – (US) Age-standardised education and income levels of male occupational incumbents from the 1950/1960 Census of Population were used to predict prestige (Berzofsky *et al.*, 2014). 3. **Warner's index of status characteristics** – (US) "a merger of occupation, source of income, type of house, and type of neighbourhood or dwelling area" (Gaur, 2013: 141).

	4. **Nam-Powers-Boyd scale (NPB)** – (US) measured average income and education of incumbents for each detailed occupational category in the census classification of occupations (1950/1960) (Galobardes *et al.*, 2006a). 5. **Pareek classification** – (India) nine characteristics: "caste, occupation, education, level of social participation of the head of the family, landholding, housing, farm power, material possession and total members in the family" (Gaur, 2013: 141). 6. **Gaur's socioeconomic classification** – (India) 7 variables: "education, occupation, income, expenditure, housing condition and living status" (Gaur, 2013: 141). 7. **Cattell SES scale** – (US; UK) "five definers of social status: prestige rating, intelligence quotient, annual income, years of education and occupations" (Gaur, 2013: 141). 8. **Subjective social status –** (US) Economic Ladder Question (ELQ) where participants self-classify their status on a 10-rung ladder (Adler, Epel, Castellazzo & Ickovics, 2000).
Traditional Composite	1. **Modified Kuppuswamy scale** – (urban India) "includes the education, occupation of head of the family and income from all sources per month" (Kulkarni *et al.*, 2013: 69).

1.3. Problem statement

SES is important to social policy and health interventions. There have been issues concerning the validity, reliability and cost of SES measurement, especially in LMI settings. For instance, in public health research the aim is to investigate how levels of inequality and social context variation affects health outcomes (Oakes & Rossi, 2003). More of the social context should be captured by public health SES measures than what education, occupation and income can offer. In LMICs, the wealth or asset index has been standardised to measure SES in communities and countries. Even though the wealth or asset index comprise of easily observable and countable features in a household, difficulties are experienced when carrying out research that deals with the collection of large-scale data. This involves the process being labour intensive, time consuming, expensive, and complicated when considering cultural barriers (Munyoro, 2018). Qualitative data can be used to enhance and confirm the SES that are pre-determined through surveys at neighbourhood, household or individual level. However, there has been no research done to confirm the accuracy of ascribing SES to neighbourhoods using qualitative data. Therefore, I explore an alternative SES measuring process that is quicker and more operationally useful for health intervention and policy planning. This approach is called the "Qualitative Ascription of SES (QASES)" in which contextual data are collected rapidly and observationally.

1.4. Aim

To understand the usefulness of a novel way to ascribe SES to neighbourhoods using rapidly collected qualitative data.

1.5. Objectives

1.5.1. To discuss how SES has been measured in health research.

1.5.2. To describe the efficiency of the QASES methodology used to evaluate SES versus a traditional, individual-level composite-measure survey.

1.5.3. To evaluate the accuracy of the QASES method relative to a gold standard measure (PC0).

1.5.4. To make pragmatic recommendations about the use of the QASES method.

1.6. Overview of Chapters

From the introduction and literature review, the structure of the thesis follows a methods chapter in which the instrumentations and procedures of both comparative methods (QASES and PC0) are described. The findings chapter entails the analyses of comparing the QASES and PC0 methods and data. Firstly, I analyse the efficiency of QASES by comparing the research labour, time and costs of QASES to a standard SES survey at individual-level using hypothetical scenarios. Secondly, I analyse the accuracy of QASES by comparing the QASES data to the PC0 wealth index data using descriptive and correlation statistics. In the discussion and conclusion chapter the results will be discussed in conjunction with relevant literature, followed by the main lessons learned, the strengths and limitations of the study, recommendations and an overall reflection of the results.

Chapter 2: Method

2.1. Introduction

This study draws on secondary data collected during the HPTN 071 (PopART) trial in the Western Cape of South Africa. Specifically, data that were collected during the Broad Brush Survey (BBS) research in 2013, where SES scores were ascribed to the study communities, and data from the population cohort (PC) at baseline in 2014, where SES variables were incorporated in a survey implemented at individual level in the study communities. I organised this chapter to provide context on the circumstances under which the data collection tools (PC0 and QASES) were administered (e.g., the research staff, their training and time to conduct the study). I develop and explain each measurement process to speak to the overarching research question on the efficiency and accuracy of the new proposed measure of SES, which is the QASES scale used during BBS data collection. This accounts for instrument validity during data collection and how/if measurement errors were minimised (Salkind, 2010). Finally, I provide the data analysis plan for my study using the secondary data by focusing on how the data were explored and which statistical tests were employed.

2.2. Overview of the research design

The research design is comparative and exploratory (Salkind, 2010). I placed emphasis on the initial data exploration while using methods that are wide-ranging to develop a deeper understanding of the data, producing new hypotheses, and to identify patterns in the data (Salkind, 2010). The goal of developing a deep understanding of the data is to examine the processes that can produce such data (Salkind, 2010). I explored the efficiency of QASES compared to an individual-level SES survey to determine if QASES is quicker and more operationally useful. I did this by conducting modelled experiments of applying the two methods to hypothetical scenarios of a small study sample and a large study sample. I aimed to determine the resource intensity (labour, costs and time) of both processes when collecting data. In addition, I also determined the accuracy of QASES by correlating and comparing QASES to an existing standard measurement of SES (PC0) making this a descriptive correlation design (Walker, 2005). The use of a comparative and exploratory research design allowed me to delve into the processes of collecting SES data, how to manage and capture the data, and testing

the strength of association between the two datasets. The validity of the new measurement tool (QASES) was determined by comparing it to an already valid measurement tool (PC0).

2.3. Intended explorations

I sought to explore whether data collection would be quicker and easier when using QASES, compared to the standard way of SES data collection. With the efficiency analysis I compared the costs, time and labour intensity of the two methods. I do this by interrogating the types of research activities required to collect SES data. For QASES this would be observational and qualitative activities (group discussions and interviews) that inform the ascription of SES. For the standard SES survey this entails a questionnaire that are conducted with individuals from randomly selected households in a community.

I also explored how well the two different measures of SES correlate with each other. I sought to do this by looking at various factors that influence the strength of the correlations. These factors are the variables of the QASES and PC0 scales. The QASES included sub-scales on housing, assets, and community outlook. Firstly, I wanted to explore whether the QASES sub-scales are sensible compared to each other. This was done by gaining a sense of whether the QASES sub-scales are internally coherent to each other before testing its correlation to PC0. Secondly, I explored how QASES performs compared to PC0 overall. I did this by investigating how QASES total performs compared to PC0. The QASES total scores per sampling zone was attributed by adding the QASES sub-scale scores for each zone. I also interrogated how each QASES sub-scale performs compared to PC0. Additionally, I looked at how different combinations of QASES sub-scales performs compared to PC0. The outcome was to identify which QASES sub-scales in combination fit better to PC0 scores. Finally, I investigated how QASES transformed scores match with PC transformed scores. This was done to see whether weighting of QASES scores or similar mathematical transformations can improve the scale accuracy. The hypotheses that follow were attempted to address these different explorations by means of statistical analyses.

2.4. Hypotheses

2.4.1. Exploring the efficiency of QASES compared to a standard SES survey

- H_0: There is no significant difference in the labour requirements between QASES and an SES survey.
- H_1: QASES requires less labour than an individual-level SES survey.
- H_0: There is no significant difference in the time required to collect data for QASES and an individual-level SES survey.
- H_1: QASES requires less time to collect data when compared to an individual-level SES survey.
- H_0: There is no significant difference in the costs for QASES and an individual-level SES survey.
- H_1: QASES is less expensive to use than an SES survey.

2.4.2. Testing the correlations of individual QASES variables

- $H_{0:}$ There is no significant correlation between QASES housing and QASES assets
- H_1: There is a significant correlation between QASES housing and QASES assets
- H_0: There is no significant correlation between QASES housing and QASES community outlook
- H_1: There is a significant correlation between QASES housing and QASES community outlook
- H_0: There is no significant correlation between QASES assets and QASES community outlook
- H_1: There is a significant correlation between QASES assets and QASES community outlook

2.4.3. Overall hypothesis

- H_0: There is no significant association between the QASES- and PC0 SES scales.
- H_1: There is a significant/statistical association between the QASES- and PC0 SES scales.

2.4.4. Total QASES and PC0

- H_0: There is no significant association between total QASES and PC0 scores.
- H_1: There is a significant/statistical association between the total QASES and PC0 scores.

2.4.5. Individual QASES sub-scales to PC0 total

- H_0: There is no significant association between QASES sub-scale of *housing* and PC0 total scores.
- H_1: There is a significant/statistical association between the QASES sub-scale of *housing* and PC0 total scores.
- H_0: There is no significant association between QASES sub-scale of *assets* and PC0 total scores.
- H_1: There is a significant/statistical association between the QASES sub-scale of *assets* and PC0 total scores.
- H_0: There is no significant association between QASES sub-scale of *community outlook* and PC0 total scores.
- H_1: There is a significant/statistical association between the QASES sub-scale of *community outlook* and PC0 total scores.

2.4.6. Combined QASES sub-scales to PC0 total

- H_0: There is no significant association between combined QASES sub-scales (*housing* and *assets*) and PC0 total scores
- H_1: There is a significant/statistical association between combined QASES sub-scales (*housing* and *assets*) and PC0 total scores
- H_0: There is no significant association between combined QASES sub-scales (*housing* and *community outlook*) and PC0 total scores
- H_1: There is a significant/statistical association between combined QASES sub-scales (*housing* and *community outlook*) and PC0 total scores
- H_0: There is no significant association between combined QASES sub-scales (*assets* and *community outlook*) and PC0 total scores

- H_1: There is a significant/statistical association between combined QASES sub-scales (*assets* and *community outlook*) and PC0 total scores

2.4.7. Wilcoxon Rank Sum Test: Transformed QASES to transformed PC0 variables

- H_0: The median of the distribution between transformed QASES total and transformed PC0 SES is equal to 0.
- H_1: There is a significant difference in the median scores between transformed QASES total and transformed PC0 SES.

2.5. Study communities

In South Africa, the HPTN 071 (PopART) study communities are located in the Cape Winelands and Cape Metropolitan Districts of the Western Cape province. These nine study communities were identified in discussion with representatives of provincial and local Department of Health and eminent advocacy organisations such as the Treatment Action Campaign, which enabled authorisation to work in these specific government health facilities (Simwinga, Bond, Makola, Hoddinott, Belemu, White, Shanaube, Seeley & Moore, 2016). These nine communities are geographically distinct to allow for the randomisation of activities, but they are also comparable regarding demographic populations and disease profiles (Hayes, Ayles, Beyers, Sabapathy, *et al.*, 2014). Their inclusion was also determined by "the health system capacity to deliver the PopART intervention and avoidance of research burden caused by participation in multiple studies" (Simwinga *et al.*, 2016:195). The primary outcome measure of HIV incidence was measured in a population cohort (PC). In the PC in South Africa, approximately 20,000 individuals participated in an interviewer-administered survey in 2014. The questionnaire in this survey included a gold-standard measure of household socio-economic status, which is used in this study.

2.6. Sample

In Figure 2, the clusters for the HPTN 071 (PopART) study area consist of nine study communities. The study communities in the colour-coded triplet clusters are grouped together on the basis of geographical location and HIV prevalence (Hayes *et al.*, 2019). For instance, the brown

cluster are located in the Cape Winelands District; the yellow cluster falls within the broader region of Khayelitsha; and the blue cluster are communities spread across the Cape Metropolitan area.

Each study community was divided into sampling zones – 18 zones in 7 of the study communities and 19 zones in the other two study communities (n total of 164 zones). These zones were demarcated within a study community and contain residential/housing sections. The zones were used to break the communities up into manageable operation sections for study field teams. The data informing the creation of sampling zones for HPTN 071 (PopART) was obtained from the Census 2011 sampled enumeration areas (EAs). Two or three census EA's/sampled zones were used to create one population cohort sample zone (see Figure 3). For the PC, a sampling fraction of 3125 houses per study community represents the population size (N). The target number of households per sampling zone was 175 to complete the baseline (PC0) questionnaire in 2014. The same zones were used for the Broad Brush Survey (BBS) study which included data collection for the QASES process.

The SES scores allocated to each zone was compared across the two scales. Each zone therefore has two measures of SES – the gold standard measured in the individual PC questionnaire averaged across households per zone, and qualitatively ascribed through the QASES process. The selected number of zones in each community are important examples of different types of SES in South Africa and these zones fall within communities with a high burden of HIV (Hayes, Ayles, Beyers, Sabapathy, *et al.*, 2014).

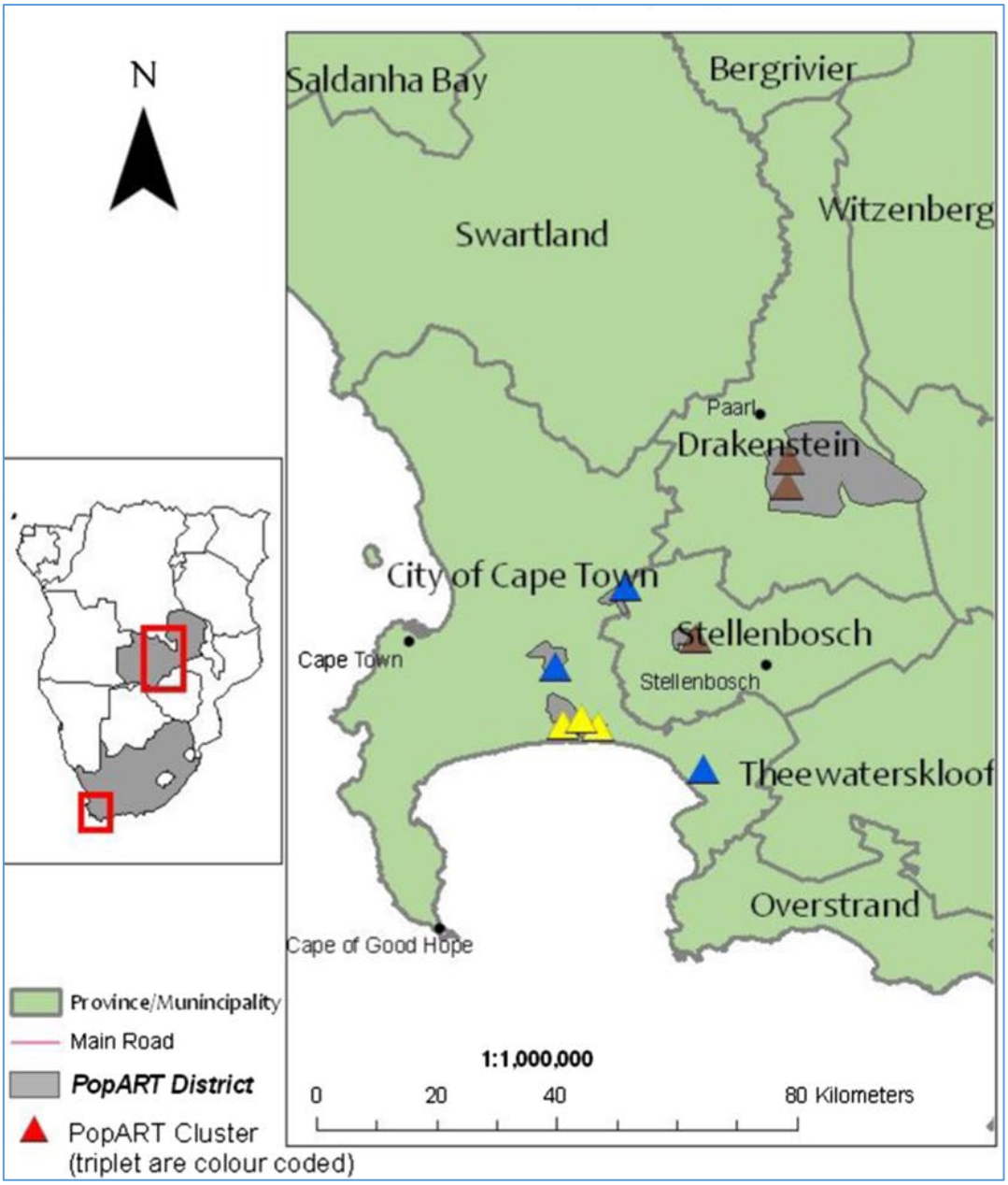


Figure 2: Map showing locations of study communities in the Cape Metropolitan and Cape Winelands Districts of the Western Cape (in Hayes et al. 2014). The communities are defined by the catchment population of a government health facility.

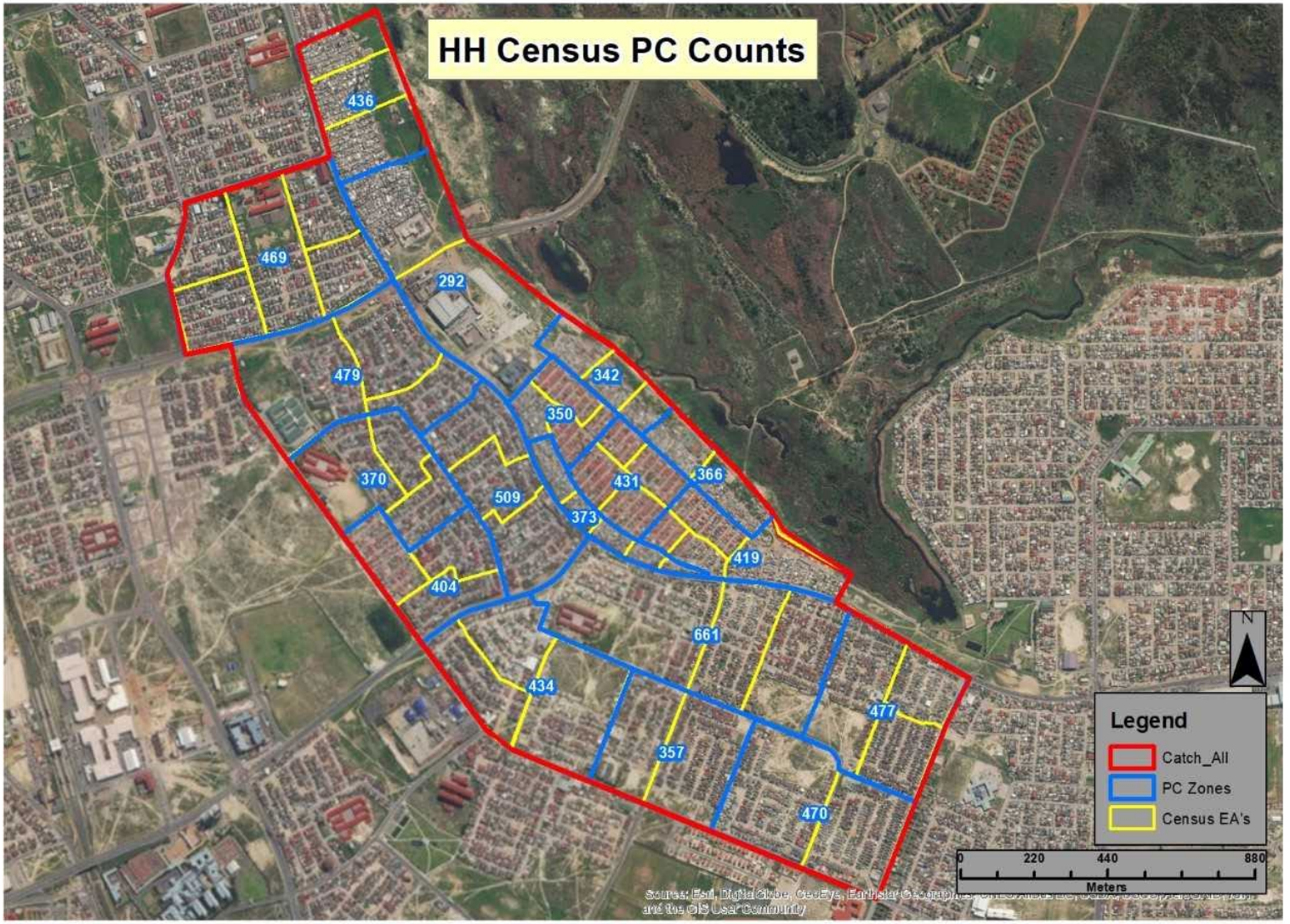


Figure 3: A study community that indicates census sampled zones and population cohort zones.

2.7. Data collection

No data was collected for this study as it is a secondary analysis of two existing data sources. I had access to SES scores for both the QASES and PC0 methods in all the sampling zones.

2.8. Characteristics of research tools (QASES and PC0)

The characteristics of the tools are described in Table 4 as an attempt to measure the variables for each method of QASES and PC0 SES, independently. The two scales are distinct in their designs and how they were implemented. I selected these methods as comparators, as they both were employed in the same study communities at around the same time period of 2013 and 2014. The PC0 questionnaire comprises of a section that incorporates a wealth index that is commonly used as a proxy indicator of SES in LMI contexts. SES is measured by adapting wealth index variables in surveys to suit the context. For instance, household assets like durable goods for consumption and production, and household characteristics like amenities, services and housing types. These indirect measures of wealth are preferred in contexts where direct variables of income have proved difficult and complex to measure. Specifically, in contexts where there is a high dependence on the informal sector and little to no fixed income. The QASES measure adopts a novel approach to collect SES information through qualitative observations of features in the communities.

Both the QASES and PC0 scales correspond through the variables employed: household characteristics, durable assets and the perception of general socio-economic outlook. The difference is that these variables are measured at different levels and perspectives. The PC0 wealth index is an objective individual-level measure where the index scores were aggregated to the neighbourhood-level in each study community. Whereas the QASES scale is a subjective tool used by the researcher to score aggregated SES scores to each neighbourhood in the study communities.

Table 4: Characteristics of each scale

	QASES scale	**PC0 SES scale**
Start of the study	2013: “Broad Brush Survey” commences in nine communities	2014: Population Cohort (0) Baseline survey conducted in nine study communities across Cape

	across Cape Town metropolitan area and Cape Winelands.	Town metropolitan area and Cape Winelands.
Type of study	Observational and cross-sectional research structured at an ordinal level.	Cross-sectional survey with continuous/interval data.
Methods	Research methods are rapid, qualitative and participatory.	Research methods are quantitative and time intensive.
Sample size	All the zones in each community are thoroughly observed, including the housing types, public transit stations, health facilities, shops, etc.	A random sample of 2,500 adults in each community (22,500 in total), aged 18-44.
Data collection	Research conducted in approximately 2-3 days for each community by three researchers exploring the areas – using a ranking sheet.	Structured surveys that are interviewer-administered by approximately 80 field staff plus supervisory staff (60-90 minutes to complete one survey) – collected over a 15-month period.
Variables/criteria	Socio-economic scale consisting of criteria: housing, assets and community outlook.	Socio-economic information is covered in the baseline survey with overarching variables of housing, assets and income/expenditures (wealth index).

2.9. The QASES method

QASES were developed from the overall BBS study. Therefore, in this section I first describe what the BBS mixed-method approach is, and how the BBS staff were selected and trained. This is followed by a description of how the BBS study enabled familiarity with the zones, how the QASES tool is used to assign SES to each zone, and the QASES processes described stepwise.

2.9.1. BBS mixed-method approach

The BBS study was formative research conducted by trained social science researchers as a "rapid pre-trial qualitative approach" and aims to inform cluster-randomised trial (CRT) interventions (Bond, Ngwenya, Murray, Ngwenya, Viljoen, Gumede, Bwalya, Mantantana, Hoddinott, Dodd, Ayles, Simwinga, Wallman & Seeley, 2018: 2). The BBS study was implemented to inform the HPTN 071 (PopART) CRT, therefore making it a pre-trial method. The purpose of BBS was to gather data both within and across complex urban communities by "systematically and rapidly observing key features of communities" (Bond *et al.*, 2018: 2). These key features integrate 4 meta-indicators that

extent towards "physical features, social organisation, networks and community identity narratives" (Bond, Chiti, Hoddinott, Reynolds, Schaap, Simuyaba, Ndubani, Viljoen, Simwinga, Fidler, Hayes, Ayles, Seeley & Team, 2016: 2). The QASES scale was designed as a way to transform insights from the BBS into a quantitative scale. QASES was designed to look at the infrastructure and population (physical, countable, features) within geographically bound locations, which are comprised in each zone of the 9 study communities (Bond *et al.*, 2018).

2.9.2. BBS staff

The BBS staff consisted of a social science lead, a social science researcher, two social science officers, and two local[2] research assistants (Bond *et al.*, 2013). The research assistants were recruited as guides during fieldwork, typically through local health committees; or they can be trained social science research assistants located, or being a resident, in the community (Bond *et al.*, 2016, 2018). The research assistants' educational background ranged from non-matriculates to those with degrees so as to offer a flexible form of employment. An effort which was strived for, is to group mixed-gendered (men and women) researchers/assistants in each community so that insights are enabled from men and women equally (see Table 5 for numbers). The implication of including a man in the group is an attempt to ensure additional security, seeing that many areas in Cape Town are considered to have a high burden of crime[3].

Racial composition coupled with language competence were other important components for researchers or assistants. In the Western Cape where HPTN 071 was implemented, it is crucial to have both Xhosa and Afrikaans speakers. There still exist a widespread mistrust of outsiders, especially of 'white' researchers (Bond *et al.*, 2018). It is paramount to be vigilant and sensitive to the composition of the communities that are typically chosen for interventions and outreach initiatives in South Africa. Nama and Swartz (2002) have noted that communities in South Africa are cognizant of the fact that

[2] Local residents from the study communities (Bond *et al.*, 2016).
[3] The homicide rate in South Africa is estimated to be six times the global average, while in the Western Cape the homicide rates were greater than the national average for both men and women (Jabar & Matzopoulos, 2017). In the province's capitol of Cape Town, the highest counts for homicide were recorded for the period between 2015 and 2016 and are spatialized in relatively impoverished sub-districts, including Khayelitsha which is one of our nine study communities (Jabar & Matzopoulos, 2017).

'white' researchers who collected data on impoverished people in the past, especially during apartheid, did not have strong intentions to improve their lives. Instead, this has resulted in academics making dubious claims on the bases of race, culture and social cohesion, which led to community members becoming increasingly cautious of this form of exploitation (Nama & Swartz, 2002). That is why it is important to establish good relations through consulting with local stakeholders to ensure that ethical research is conducted. Local concerns and customs should be considered to support the general well-being of the community (Mosavel, Simon, Van Stade & Buchbinder, 2005).

Table 5: BBS field staff according to sex for HPTN 071 (PopART) in South Africa

	Senior social scientists		**Field team social scientists**		**Local research assistants**	
Sex	Men	Women	Men	Women	Men	Women
PopART	1	1	0	2	2	0

(Adapted from Bond *et al*., 2018)

2.9.3. Training of BBS staff

Training of research staff took approximately one week and were conducted at the site office or in the community. The training was aimed at familiarising research staff with qualitative skills such as reflexivity, observing, community characteristics and layout, and writing of textual data. Additional training also included orientating researchers to use research tools (e.g., interview guides, group discussion guides, and observation tools) and allowed them to practice using the tools in sessions and teams. A significant aspect during the training was to make fieldworkers aware of the CRT and the role that they play within the CRT (Bond *et al.*, 2016, 2018). For example, so that the researchers were able to introduce the study to people when they were approached in the community or when they spoke to people.

Furthermore, the researchers had to reflect on the community entry/exit points and any relevant ethical dilemmas that they may have encountered while doing fieldwork. It was a requirement for researchers to attend Good Clinical Practice (GCP) courses so that they are informed

on ethical approaches[4] to social research. The main teams also went for training in data management and data capturing.

2.9.4. Familiarity with the zones

Together the researchers formed a foundation of what the communities are like from the qualitative BBS mixed-method data collection efforts. In Figure 4 the activities included, amongst other things, a spiral walk activity – giving an overview of the layout of the community; and in-depth interviews, group discussions and observations of communities – providing insight into the households in the different zones (Bond *et al.*, 2013). The sequence of research activities is described in Appendix C and it initially started with broader observations leading to narrowed down and structured observations in places of gathering, access points, and other spaces, like local health facilities. Furthermore, the observation periods that took place over 3-5 days, formed the necessary structure and themes of activities for the BBS since these activities are crucial for capturing the meta-indicators[5] (Bond *et al.*, 2018).

The group discussions happened simultaneously as the observations and information was gathered on (1) how different groups would describe their communities to answer the question of 'what kind of place is this?', (2) HIV mapping to identify key places in the community, (3) wealth, poverty and risk taking to tell stories and develop characters of people living in the community, (4) concept mapping through a series of activities of writing down what HIV prevention is in the community, to group all concepts that fit together in a pile, and to rank concepts in terms of importance for prevention in each community, (5) Institutional mapping where all the organisations and groups were listed that work on HIV prevention in the community, and (6) HIV timeline drew by community members of all relevant HIV service activities in the community. See Figure 4 for all the BBS activities listed. Finally, the key informant interviews were done with people identified during

[4] The teams are prepared for situations where they may experience a crime or witness a crime, for events where people ask them about health conditions associated to the trial, and for instances where people may approach them for social support and having to refer people to support services (Bond *et al.*, 2018).
[5] Infrastructure and population (physical, countable, features); Social organization (relation of people to place, choice among the options); Networks (relations of people to people, patterns of inclusion & exclusion, control of local resources); Narratives (stories about us, who is in the moral community, who belongs here).

the group discussions that have a significant role in the work towards understanding and eradicating HIV in their community (e.g., with community health workers like nurses and doctors, traditional healers and NGO workers).

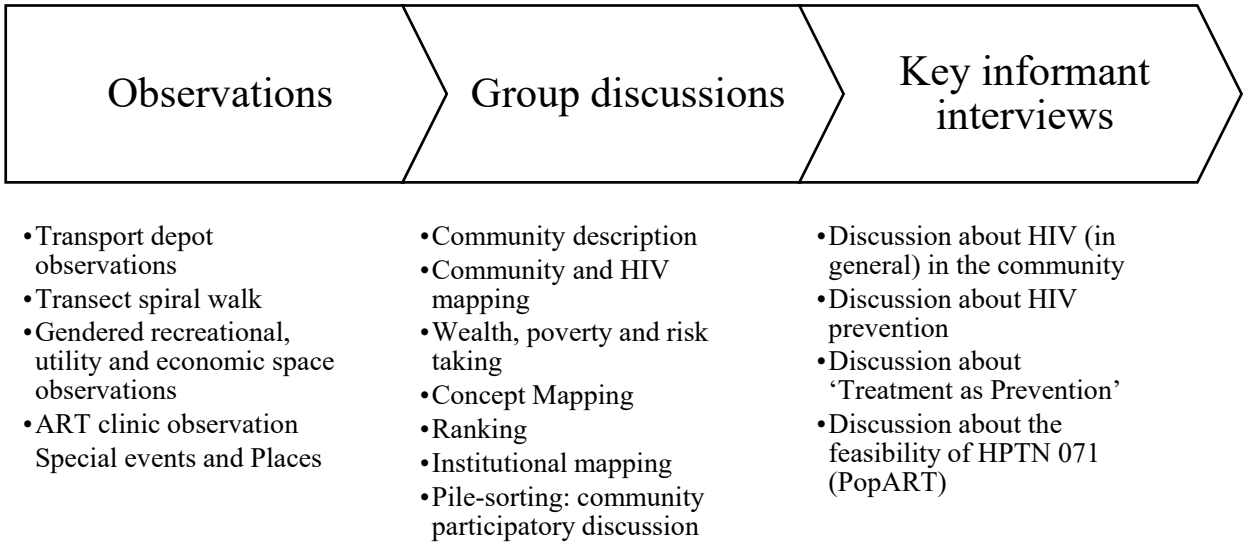


Figure 4: BBS key activities

2.9.5. Ascribing SES to each zone

After the BBS data collection was completed, three social science researchers used maps of the zones and scored them based on three socio-economic indicators: housing, assets and community outlook (see Appendix A). For the sub-scales, effort was based on the combined judgement of the researchers to determine how to score each sampling zone within a study community. The tool is ordinal in nature, which means that each number indicates an order of magnitude with "no numerical meaning beyond the order" (Garth, 2008: 3). Each sub-scale has a scoring level from 0 to 4, with 0 being very poor, 1 being poor, 2 being adequate, 3 being good and 4 being very good. All three sub-scales can be considered as subjective based on the researchers' opinion on the average in that zone. For the housing sub-scale, each zone was investigated based on previous fieldwork done during the BBS study and also observing the maps. The researchers determined that most people in a specific zone live in houses that are closest to a score X (0-4). The same can be said about assets as per

previous observations and reflections from the BBS fieldwork, an overall score was assigned to each zone. The third sub-scale of community outlook was used to rate a zone based on a spectrum ranging from hopefulness for progress to abject misery with no outlook of progression. The researchers got a chance to look at a zone at a deeper, subjective level where interviews and conversations on the street with people residing in the area might have assisted in deciding what score to give during BBS fieldwork.

2.9.6. QASES Process

For the BBS study in 2013, the QASES ascription process was only initiated after the main activities were completed[6]. The intention with the QASES scale was to indicate how homogenous or heterogeneous zones within a community is, without only looking at homogeneity/heterogeneity on an inter-community level. The QASES scale can be applied independently when conducting community health research to gain quick insight on the overall characteristics of a community and how they differ moving around in the community. From figure 5, the steps are as follows:

Step 1 was to recruit a team for BBS. The staff were one social science lead, one social science researcher, two social science officers and additional research assistants (~2) (BBS Technical Report, 2013).

Step 2 was to offer training to the research staff. This included training the core scientific staff on the BBS methodology and research tools for a period of approximately one week. From there a pilot study in one community was initiated for operational refinement of implementation processes going forward. This also guided and enabled training for the field staff to make further adjustments following the pilot study (~1 month). The pilot study was introduced to refine the researchers' ability to approach and speak to people in the community (specifically to explain the study and what information they require from the community members).

Step 3 was to collect data in the remaining study communities (~5 months) (BBS Technical Report, 2013). The BBS mixed-methods data collection (as shown in Appendix C) took 12 days per

[6] The activities form part of the entire BBS method, whereas I only focus on describing the QASES ascription process. The additional information are provided as an overall guide to BBS and what other ways might contribute to and make things easy when trying to measure SES in communities.

study community. Debriefing sessions also happened daily that helped to orientate the team on the most important findings of the day and facilitated the organisation and processing of the collected data (BBS Technical Report, 2013).

Step 4 was for researchers to assign QASES scores to each zone by inspecting the study area maps. They had to be familiarised with the QASES tool in order to understand how to ascribe SES scores. What also facilitated in ascribing individual scores (based on sub-scales), was for researchers to reflect together from their fieldwork activities during BBS (observations, group discussions and interviews). This guided researchers to give an overall score to each zone according to the QASES sub-scales.

Step 5 was to ascribe a total score to each zone based on adding the values of the sub-scales. The processes in step 4 and step 5 took approximately 2-3 days per community to complete, depending on the size of the community and the number of zones.

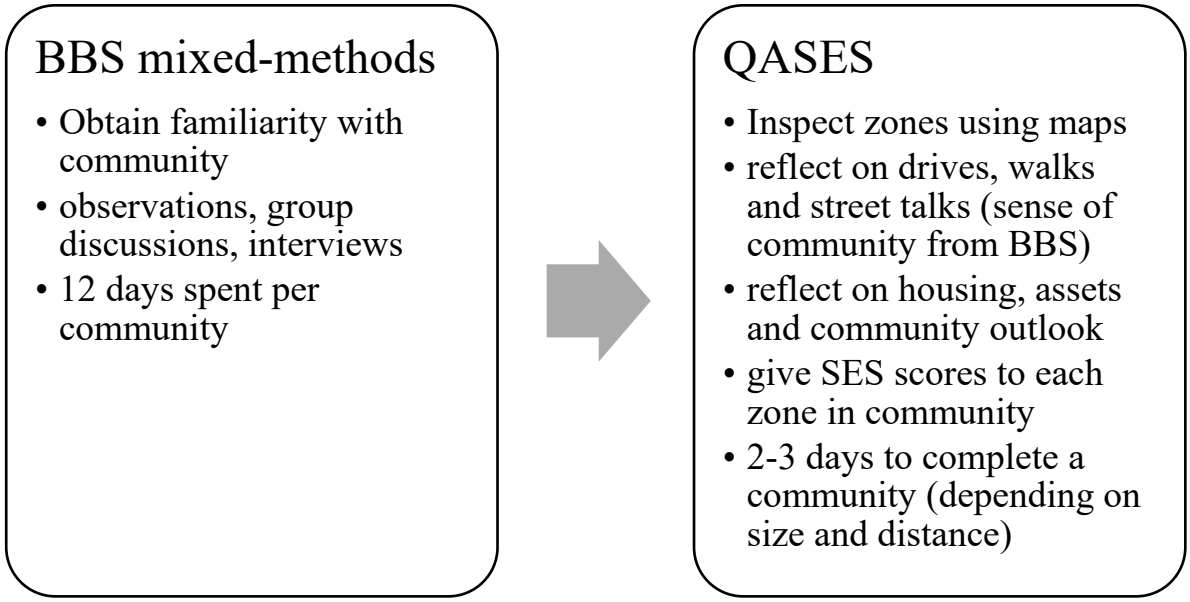


Figure 5: Sequence of QASES procedures

2.10. PC0 SES method

The data collection SES tool for PC0 is validated as the accurate gold standard SES measure. By gold standard, I refer to the way the data is collected, as well as the use of standardised socio-

economic indices in the survey, for instance the asset index/household wealth index[7]. The household wealth index expands towards income or expenditures variables, household possessions or assets, and household amenities (like type of flooring, drinking water source and toilet facilities) (Fotso & Kuate-Defo, 2005: 196). The PC0 SES measure is classified as an objective composite measure that are applied at individual level in the selected study communities. The data of investigation involves a baseline survey where an individual-level questionnaire was done with one member of the household on various topics, including an economic section which covers the person's socio-economic status (see Appendix B).

2.10.1. PC0 fieldworkers

Fieldworkers were recruited from the study communities[8]. Recruitment from communities is essential because they understand the community dynamics, they were familiar with the area, they knew how to approach people – this guided the study team to approach and reach household members (e.g., best times to visit people at home, who are generally employed/employment levels). More women were employed than men, this is because more women applied than men. It was a requirement that they had 2 years relevant experience (with matric and specifically in research) and acknowledgement of community leadership (as references). A total of 80 fieldworkers were recruited during the process of PC0 data collection. PC fieldworkers worked in pairs in the communities (approximately 8-9 fieldworkers per study community). They worked in pairs due to safety reasons and the process was also quicker to collect data by sending a minimum of four pairs to one community (depending on the community size).

2.10.2. Training of PC0 research staff

In the Population Cohort at baseline (PC0) practical training of fieldworkers entailed: (1) administering GPS devices (viewing/navigations/saving GPS coordinates), (2) the survey (working with the electronic tablet devices and understanding the questions), and (3) teach fieldworkers about

[7] Asset index and wealth index are often used interchangeably in the literature (Howe *et al.*, 2008)
[8] This is part of building strong community engagement to establish partnership between the researcher/fieldworker, participants and communities and to ensure that ethical responsibility are enacted in the study communities (Hayes & Fidler, 2015).

health studies, (including the CRT, HIV, the context and research aim). The training of fieldworkers took 1-2 weeks. Additionally, all study staff received Good Clinical Practice (GCP)/research ethics training before the intervention implementation commenced (Hayes & Fidler, 2015). Many employees came from different research backgrounds, like collecting data for StatsSA or environmental sectors, therefore, they had to receive training on public health research skills, specifically how to understand HIV within the communities that they are working in. This is to ensure that they are capable of doing the following when recruiting and interacting with participants (Hayes & Fidler, 2015: 52):

- To thoroughly explain the study visit schedule and procedural requirements during the informed consent process.
- To thoroughly explain the importance of their participation that contributes to the study's overall success.
- To be able to collect locator information at the study enrolment visit, and active review and updating of this information at each follow-up visit.
- To increase awareness about HIV/AIDS through regular communication with the study community at large and explain the purpose of HIV prevention research and the importance of completing research study visits.

2.10.3. PC0 SES questionnaire section

The section of importance in the PC0 questionnaire is titled 'economic activity and food security' (Appendix B). This section is also divided into sub-sections with questions focusing on housing, assets, food security, occupation and income/expenditures. In the section on housing, questions were asked about what type of structure and flooring the house has, how many rooms there are, and what type of municipal services they receive. In the assets section, the participant had to identify which durable assets they have in their home out of a list that includes a cell phone, motorcycle/scooter, car, bicycle, electricity, radio, TV, fridge, CD player, laptop, and stereo/music playing device. The sections on income/expenditures and occupation focus on what the main sources of income are, how much they spend on commodities/necessities, if the participant is employed formally or informally, and how much dependence there are on government grants. Food security

questions entail whether the participants had received relief or free food, had to cut down on certain foods, and had to skip any meals during the last twelve months (See Appendix B). Questions were binary (yes/no), categorical (race, sex, age group, and educational level), numeric (exact age; number of rooms in your house), and ordinal (level of agreement/Likert scale).

2.10.4. PC0 processes

The processes for PC0 took place in four phases which is summarised in Figure 6.

2.10.4.1. Sampling of Population Cohort households

The PC zones were selected based on the Census 2011 samples Enumeration Areas (EAs). The total number of households in each PC zone were approximately 350. Each study community underwent a household enumeration process where GPS coordinates of all houses in a zone were captured, all individuals in a household were listed (based on age and sex), and they were asked guiding questions regarding their household (e.g., sharing of amenities). About 38-40 fieldworkers were required to collect the data of the household enumeration process. The enumeration process informed which households are eligible for the PC. To get a representative sample for the PC study, organised stratified sampling was applied to randomly select households for the PC (groups of households in geographical proximity that are labelled as zones n=175). This means that approximately175 households per zone, or 3125 households per study community, were eligible to partake in the PC study. Successful recruitment is not necessarily dependent on meeting eligibility criteria informed by household enumeration when collecting SES data. I simply included the process that was used to create a sample of households for the PC study.

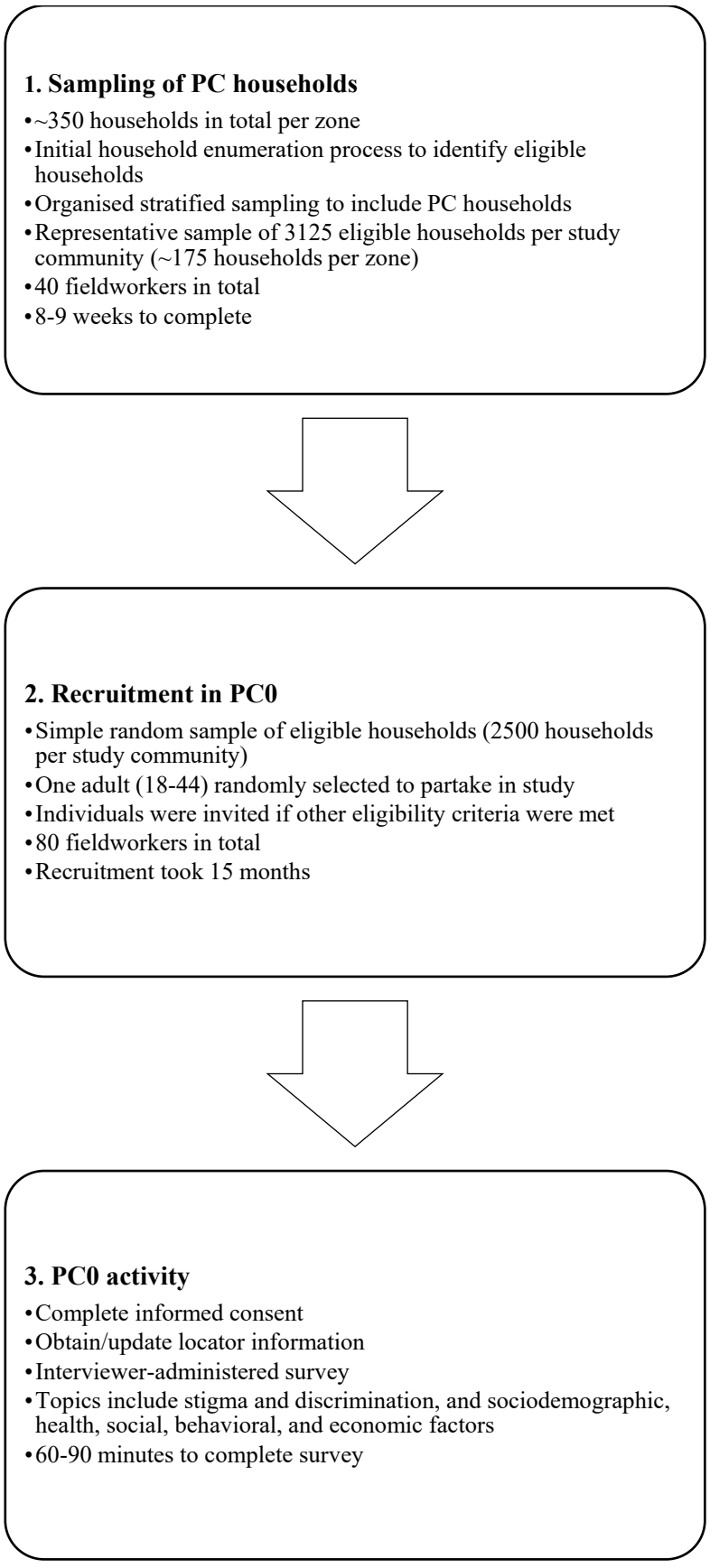


Figure 6: PC0 processes in sequence

2.10.4.2. Recruitment in PC0

PC0 data collection required 80 fieldworkers when recruitment of participants commenced. After a simple random sample of houses were selected (~2500 households per community) of the eligible households (from organised stratified sample), one adult per household (18-44 years of age) was randomly selected from the household list to be included in the PC. The eligible age range meant that individuals 18 years and older can participate without parental consent, and adults younger than 45 years are projected to experience a measurable change in HIV incidence due to the study intervention (Hayes & Fidler, 2015). The individuals that are selected were only invited to participate in the PC if they met the other eligibility criteria. Recruitment of participants took 15 months for PC0.

2.10.4.2.1. Population Cohort Inclusion Criteria

- Aged between 18–44 years
- Willing and capable to provide informed consent
- Living within the catchment/study area of a selected local health unit and expecting to reside there for the duration of PC (3 years)
- Living in a randomly selected household

2.10.4.2.2. Population Cohort Exclusion Criteria

- At the time of PC, were enrolled in another HIV treatment, prevention, or Pre-exposure prophylaxis (PrEP) study
- Enrolled in an HIV vaccine study at the time of PC and prior to PC.
- From the investigator's opinion, anything that impeded informed consent, caused safety issues for participation, complicated interpretation of study outcome data, or interfered with achieving the study objectives.

2.10.4.3. PC0 activity procedures

I want to iterate that the procedural steps to conduct the PC0 study encompass what was done for the HPTN 071 (PopART) CRT, which are specifically relevant for the HIV incidence estimate. I only state the steps that was taken to ascribe the neighbourhood SES in the PC0 zones, by excluding the longer and overall process of collecting the survey data (e.g., the random selection of an eligible household member to be included in the PC). The relevant steps for ascribing SES were as follows:

Step 1 was to randomly select households from each neighbourhood/community zone to be included in the PC0. The number of households per community zone were approximately 175.

Step 2 was to invite one eligible adult from each of the randomly selected households to participate in the PC0. This was done by the PC fieldworkers upon participant recruitment.

Step 3 was for the participant to receive informed consent on the PC0 study. The fieldworker had to explain to the participant, in their preferred language, what the risks/benefits are of participating in the study.

Step 4 was to update and obtain the participant's locator information using a GPS. This was to ensure that geographical coordinates are available for each household that have been recruited. It was also necessary to have the locator information in order to track which households needed to be reattempted when the first visit was not successful.

Step 5 was to complete the interviewer-administered survey with the participant focusing on "topics of stigma and discrimination, and socio-demographic, health, social, behavioural and economic factors" (Hayes & Fidler, 2015: 51). The survey took about 60-90 minutes to complete.

Step 6 was to transform the PC0 data from a collection of the survey answers, some of which are true/false, some of which are Likert scales etc., to continuous variables.

Step 7 was to generate an asset/wealth index from principal components analysis (PCA) of data on the following assets: household construction materials, water source, type of floor, energy source for home, sanitation facilities, source of income, government grants, food security and ownership of: bicycle, cell phone, motorcycle/scooter, car, household electricity, fridge/freezer, television, radio, computer, and/or music playing device.

Step 8 was to run principal components analysis to derive weights separately for each HPTN 071 (PopART) country, i.e., South Africa and Zambia. Then tertiles were created, before combining the data across the countries.

Step 9 was to aggregate the PC0 household-level data to zone level which were adjusted as normalised data. Each study community had demarcated zones where the household-level data were collected. The wealth index items, together with the additional economic activity and food security

questions in the questionnaire section, were summarised for all the households in a zone by determining averages of all the completed questionnaires.

2.10.4.4. PC0 data quality improvement

In terms of resource intensity, quality control and improvement were implemented throughout the PC processes before data collection commenced. This is to determine where to start working, how much time to spend in each community/zone or sending a driver in to assist (safety or size of the community) – i.e., how to approach each area individually. When data collection for PC0 were in progress, fieldworkers were not always successful to reach all the randomly selected households in a community zone. Therefore, multiple attempts were made which were determined by going at different times (afternoons or weekends), returning there again, and asking the neighbours about who lives in the specific household and if they are currently residing there. The attempts were monitored and sometimes a quota was reached for attempts made (for instance, a quota of 3 attempts before excluding a household). The success rate for reaching all randomly selected households were approximately 90%. Each case of unsuccessful attempt was investigated individually before excluding them (case-by-case approach). Ultimately, this explains that not all randomly selected households were reached and included in PC. This resulted in bias when only a few households (in a zone) completed the survey instead of all randomly selected households.

2.11. Data analysis

The data was analysed using SPSS version 26. The QASES data are ordinal, meaning that it was ranked or ordered linearly. The PC0 data were converted to an interval scale using a standardized norming conversion. Interval data are also referred to as integer data and entails ordering and distance measurement, of which there is meaning between the data points (Allen & Seaman, 2007; Miller & Yang, 1997). Descriptive statistics were run on the data, along with the non-parametric tests of Spearman correlation (rho) (significance level of 0.01) and Wilcoxon Rank Sum Test (significance level of 0.01). Spearman rank correlation (rho) was used to explore the relationship between the two different forms of measurement observed as the socio-economic status variables. Wilcoxon Rank Sum

Test was applied to the variables of QASES total and PC0 SES in order to identify the magnitude of the observed differences between the two measures (Field, 2009).

2.11.1. Data cleaning

I received the datasets of both QASES and PC0 already captured in Microsoft excel format. My co-supervisor (Dr Dunbar) was responsible for sorting/cleaning the two datasets. We received clearance to use the PC0 data from HPTN 071 (PopART) (see attached in Appendix E).

2.11.1.1. Alignment of PC and QASES data

After inspecting the data, my supervisors and I realised that the BBS QASES data were not aligned with the PC0 data. When we investigated the maps for both studies showing the BBS zones and PC zones, we came to learn that the demarcation of zones was not completely contiguous. For instance, even though both studies copied their zones from StatsSA sampling enumeration areas (EAs), the numbering of the zones differed for PC and BBS. Additionally, some individual PC zones comprised of 2 BBS zones combined (see Appendix F of reworked maps to align BBS with PC). My co-supervisor and I had to rectify the maps so that the PC and BBS scores can align in the dataset. Due to the process of combining and matching zones for improved alignment, the number of zones decreased from 164 to 145.

2.11.1.2. Details of QASES and PC0 data

The data were captured in excel format containing the 9 study communities with their individual zones (145 zones in total). The variables entail the QASES individual sub-scales with their ordinal values (e.g., 0=very poor to 4=very good); the QASES total variables which is the sum of the individual sub-scales (up to the value of 12); and the PC0 SES values captured as continuous (fractional) values on an interval scale (-3 – 3).

2.11.1.3. Outliers

Part of the data cleaning process upon initial investigation, was to create the graphs and correlations between the data to determine outliers. From the scatter plot of QASES total and PC0 SES, outliers were detected. I ran a Stem and Leaf Plot of the PC0 SES data to determine any extreme outliers (see Appendix G). I removed three values from the entire dataset reducing the amount of data points from 145 to 142. I initially applied the statistical analysis to the unaltered data and from there

on the outlier values were excluded from the continuing analyses to reduce bias. These 3 extreme values (zones 28, 29 and 128) were detected in the PC0 data. There was a great disjunction between the scores given for these zones compared to scores given to adjacent zones. One possible explanation for significant differences between PC0 and QASES is that the PC0 measure is somehow inaccurate. For example, if the PC average were based off a small number of surveys. However, in this case all three zones had samples of more than 100 surveys so this is unlikely to be the cause of the outliers. A possible reason for the discrepancy between the PC0 and QASES scores here was that the PC and BBS zones may not have overlapped perfectly. However, the existence of outliers may also be a limitation of the QASES method discussed in section 4.7. I concluded that such outliers should be further investigated in future applications of the QASES process.

2.11.2. Initial data exploration

2.11.2.1. Descriptive statistics

I started by running descriptive statistics to describe the range, average, most common category and distribution of each variable (housing, assets, community outlook, total, relative total, and PC0 wealth index). I made simple bar graphs of the QASES data and PC0 data to illustrate the number of points at each level of frequency (very poor, poor, acceptable, good, very good); the total scores of the combined variables (housing, assets and community outlook); and the PC0 SES distributions (Simpson, 2015).

2.11.2.2. Simple scatter plots and Box-and-Whisker plots

I made scatter plots to see what the associations are between the QASES and PC data. I did so for the QASES total scores, and combined QASES sub-scale scores versus the PC0 SES scores. Besides checking for a relationship between the variables, scatter plots also show what type of relationship it is and whether any data points are distinctly different from others (Field, 2009). The outliers were detected from the scatter plots for the PC0 SES data to identify how they influence the distribution. From there, I checked for specific outliers in the PC0 SES data by running descriptive statistics and creating a stem and leaf plot to detect the extreme values (see Appendix G).

I also made Box-and-Whisker plots of the individual QASES sub-scales compared to PC0. This was an attempt to visually illustrate “the range, distribution symmetry, and central tendency of a

distribution in order to illustrate the variability and the concentration of values within a distribution" (Lewandowski & Bolt, 2010). The centre and dispersion of data for the PC0 variable for each QASES sub-scale category are graphically represented with the five-number summary, including the minimum value, 1st (lower) quartile (Q1), median, 3rd (upper) quartile (Q3), and the maximum value. For each box plot, outliers are also indicated.

2.11.3. Non-parametric tests

Non-parametric statistics are used when data have violated parametric assumptions, like non-normally distributed data (Field, 2009). In this case the distributions of the graphs and scatter plots were normal. However, the data for QASES and PC0 are different, as QASES data are ordinal and PC0 data are interval. According to Field (2009), fewer assumptions are made of the data when non-parametric tests are applied. Bar graphs/histograms were used to show the range on the scale and distribution of each individual variable. Scatter plots and box-and-whisker plots were used to show the range on the scales and distributions between the method variables.

2.11.3.1. Spearman's correlation coefficient

The non-parametric statistic of Spearman was used to correlate the strength of the associations between the variables. I ran the correlation tests between the variables of QASES (as individual variables and total variables) and PC0 SES. For Spearman's rho, the data are ranked. For instance, the lowest score is detected and given a rank of 1, then the next highest score is detected and given a rank of 2, and so forth (Field, 2009: 540). This process means that the lowest scores are represented by small ranks and the highest scores are represented by large ranks. Ultimately, the analysis is carried out on the ranks rather than the actual data (Field, 2009). I completed this process computationally using SPSS version 26.

2.11.3.2. Cross-tabulation and Wilcoxon Rank Sum Test

The data for both QASES and PC0 were transformed (see Appendix H for how data was transformed). I did this to perform the Wilcoxon Rank Sum Test, together with creating an aggregate graph of the transformed data (for both QASES and PC0 SES). I wanted to determine how well the QASES scores match with the PC0 SES scores when the data for both were organised in an ordinal manner. The purpose was to count how many times QASES scored similarly to PC0 SES. Cross-

tabulation was run between the two transformed scales, to see the numerical representation of how many times QASES and PC0 SES scores match (in total) (see Appendix I). Furthermore, a visual representation in the form of an aggregate graph was made that indicates the amount of values per block similarly scored between the two scales in the colour red (the darker the colour red, the more similar scores there are).

The non-parametric Wilcoxon rank sum test was done to identify the magnitude of the observed differences between the transformed QASES and PC0 SES scores. The test compares two paired groups (QASES and PC0 SES). The observations were sorted according to the test variable (PC0) and ranks were assigned to each observation. SPSS was used to conduct this test. By doing the Wilcoxon test, I determined how big the difference in scores are and approximately how close does QASES scores get to PC0 scores. The aggregate graph of the matched QASES and PC0 SES scores emphasises the differences and approximates visually.

2.11.4. Specific significance tests of Spearman's correlations

Spearman's correlation coefficient was used to test the correlations between the BBS QASES and PC0 measures of SES. The specific tests of correlations were to reject or accept the two-tailed hypotheses, meaning that the hypotheses do not suggest the direction of the relationship making them non-directional (Field, 2009). The purpose was to understand how well the strength of associations between the two different measures were by performing various correlations. This gave an idea of what variables influence the associations to be stronger or weaker.

2.12. Ethical considerations and clearance

2.12.1. Connection to HPTN 071 (PopART)

This study contributes to the research outputs and considerations of HPTN 071 (PopART) of reducing HIV infection in high-incidence communities. The impact of a person's and a community's SES on their health is an important factor in determining HIV incidence. This is due to low SES being associated with high HIV risk and poor access to HIV services. To understand why and how SES

circumstances would shape high levels of HIV infection in these areas, provides important contextual data to the PopART intervention.

2.12.2. Risk/ benefit

The study is low risk seeing that no additional data were collected on the PC participants. No personal information of the participants is made available to this particular study except for the zone (approximately 350 households) in which the participants reside. This, however, is not revealed in the study findings, results or discussion. There is also no involvement of any data that is indicative of the HIV status of participants. Knowledge gained from this study is to create and develop new ways of understanding the SES indicators of communities without having to involve participants to share in-depth information about themselves. This can ultimately lead to research that is less burdensome on research participants in terms of time required from them.

2.12.3. Autonomy

The participants were randomly selected in the households which were also sampled randomly. Thus, the selected participants were given all the relevant information about the study and were open to participate or withdraw at any stage during the process. No consent was necessary for QASES as data were collected through the BBS process which were mainly observational and all interview participants consented to this use of the data.

2.12.4. Procedural

2.12.4.1. Data storage

Data for PC0 (initial survey) and BBS is stored at the Desmond Tutu TB Centre (DTTC), Stellenbosch University. The data was handled with care seeing that the information comes from a vulnerable population sample based on general low SES settings and high burden of HIV.

2.12.4.2. Anonymity and Confidentiality

The confidentiality of the participants is protected through giving each one a number. All information is labelled with this number only and personal information (name, address, phone number) of participants are protected by the research staff. None of the personal information are used in any publication of information about this study.

2.12.4.3. Informed consent

This is a sub-study of data obtained in PC0 of which informed consent has been granted making it possible to use the data for future studies within PopART (Appendix D).

2.12.5. Ethical clearance

I received ethical clearance from the Health Research and Ethics Committee, Stellenbosch University on 16 April 2019 (Appendix K).

Chapter 3: Findings

3.1. Introduction

In section one of the chapter, I describe the efficiency of using the QASES method compared to a standard SES method in the form of an individual-level survey aggregated to neighbourhood level. I present two hypothetical examples of the resources it would require measuring SES using the QASES versus the standard SES survey. They are to measure SES (1) for eight neighbourhoods in one community and (2) for 100 communities in a city. I determined the resource intensity of cost of employment and time required to complete data collection using both methods.

Section two is structured to look at the accuracy of QASES: (a) correlations of SES measured in the two ways using Spearman's rho – statistically indicating the strength of association between the two measures of SES. And (b) transforming the QASES SES measure in a variety of ways to explore the possibilities for improving its validity relative to the gold standard using Wilcoxon Rank Sum Test. I visually illustrate and discuss the findings from the descriptive and bi-variate correlations using bar graphs, scatter plots and box-and-whisker plots. From the bar graphs I explain the range of the variables (lowest and highest values); average and most common category. This is used to look at the distributions of each variable. The scatter plots and box-and-whisker plots visually illustrate the distributions and associations between the variables.

Statistical significance is set at 0.01 which is a 99% confidence interval. This means that there is 99% confidence that the data are generalizable to the wider population and therefore similar results will be found when applying this method to similar SES contexts. Wilcoxon Rank Sum Test statistically examines the magnitude of the observed differences between the two samples at a statistical significance of 0.01 or 99% confidence interval.

3.2. Efficiency of the QASES method compared to a standard SES survey at individual-level

I conducted two hypothetical scenarios where I compared the efficiency of measuring SES in a small sample setting and a large sample setting using QASES and a standard SES survey at individual-level. For these hypothetical scenarios I used the term community to describe a broader

geographical area that includes approximately 8-15 residential neighbourhoods. Each of these neighbourhoods has approximately 250 households/1250 residents. In these two scenarios I determined the labour force and costs of applying the two different methods while working out the activities and time required to complete each study. I focused on the amount of staff required to conduct both studies and determining how much they would cost based on their qualification requirements and salary payment per hour. The idea behind this exploration is to determine the major differences in terms of resource allocation for each method. There are additional costs when applying both methods that are not accounted for in these scenarios. For instance, training days of staff, data collection planning, traveling costs (e.g., car rentals and petrol) and equipment requirements (e.g., printing of papers, electronic surveys or GPS's) were excluded, as I only explore the data collection processes. For the purposes of this comparison, it is safe to assume that the larger staffing component necessary for an individual-level data collection will mean that these operational costs are also greater. Therefore, any comparison presented here is an underrepresentation of the true cost saving in the QASES process.

3.2.1. Hypothetical experiments of applying the QASES and PC0 methods

3.2.1.1. Example 1: Small study sample in the community of Smithson

A clinic manager plans an HIV-tent testing campaign in the community of Smithson which comprise of 8 neighbourhoods. The clinic manager wants to know in which of the 8 neighbourhoods the lowest SES is so that HIV-testing tents can be set up in those neighbourhoods. The Department of Health sent out a notice to research tenders for the study to be conducted. They eventually end up with two possible tenders: Option A, they do individual data collection from a random sample of households in the 8 neighbourhoods using a standard SES survey. Option B, they do QASES in the 8 neighbourhoods. The options are weighed against each other regarding the labour intensity, costs and required time to collect data.

3.2.1.1.1. Option A: standard SES survey in Smithson's 8 neighbourhoods

For option A, I estimated the resource requirements for conducting standard SES surveys as shown in Table 6. To expand on the study area, the community of Smithson has 8 neighbourhoods

comprising of approximately 2000 households in total. Each neighbourhood includes approximately 250 households. When applying the standard SES survey at individual-level, a random sample of households need to be determined in each neighbourhood. An approximate of 200 households were randomly selected out of all the neighbourhoods that act as a representative sample (25 households per neighbourhood). That reflects about 10% of the total households in each neighbourhood. Therefore, in Smithson 200 households had to be visited by fieldworkers to complete an SES survey.

The questionnaire to measure SES at individual level includes approximately 20 items, which ask a range of questions pertaining to education, occupation and income (in the form of a wealth index). For the completed number of individual surveys to be aggregated to each neighbourhood zone, a random sample of households must be selected that are representative of the households in the neighbourhood (approximately 25 random households). I estimated that a fieldwork pair can successfully complete approximately 8-9 surveys per day assuming that recruitment, plus consent, plus survey completion would take approximately 45 minutes per participant within an 8-hour working day (1 hour for break and lunch). The rate would be that one pair of fieldworkers should take 3 days to complete 25 surveys per neighbourhood. Therefore, with 2 pairs of fieldworkers, data collection in the neighbourhoods could be completed in approximately 12 days (see Table 6). This excludes the time lost to find an eligible participant within an economically active age range at home when visiting the randomly selected household and having to return at a different time.

The qualification requirements of the fieldworkers are that they completed secondary schooling with some background in research data collection. I estimated that these fieldworkers earn R52.36 per hour (see Appendix L of Stellenbosch University's 2020 base remuneration post-level bands; Table 7). The team of 2 pairs of fieldworkers have to be supervised by a more senior researcher, specifically concerning the data capturing, checking and reporting. A senior research supervisor earns an estimated R228.46 per hour (Table 7). The additional costs and time required to process the SES results in the community are dependent on the instrument of data collection (e.g., using a paper survey or electronic device). I do not include such estimates because I only focus on the time and labour required while in the field collecting data. It is, however, a quicker process to use

electronic devices that can transfer and store the data once the survey is completed, compared to a researcher who has to manually capture all survey answers into a dataset. In Table 6 I indicated the total costs of labour and required time to collect 200 surveys in 8 neighbourhoods.

3.2.1.1.2. Option B: QASES in Smithson's 8 neighbourhoods

In option B I applied QASES in the 8 neighbourhoods in Smithson. QASES were adapted to include the BBS observational activities, group discussions and interviews. The observations in each neighbourhood inform the levels of SES according to housing, community outlook and assets. The observations are as follows: observe the neighbourhood structure in terms of (1) entry-exit points, (2) economic places, markets and services (e.g. gendered spaces like bars, hair salons/barbers), (3) do a transect walk from local clinic/health-care centre/hospital to observe housing types, infrastructure, available facilities and movement of people, (4) visit transport depots to observe movement (5) approach people on the streets, while doing observations, and have short conversations with them about what it is like living in their neighbourhood. The topics are regarding their understanding of their neighbourhood's socio-economic class structures, specifically: social mobility; access to public/private transport; access to services; types of housing; aesthetics of neighbourhood; safety and crime.

Group discussions involve government officials (social development and other expert departments/organisations), old men, young men, old women, young women. These community members are reached in the streets while doing observations and at the related government/organisation facilities (5 group discussions). Key informant interviews are conducted with local stakeholders who are knowledgeable in the field of socio-economic profiling in the community, like government officials (e.g., social development) and members of NGO's (~4 interviews). The interview and group activities focus on discussions of: general education levels; employment rates; social mobility; access to services; socio-economic class distributions; crime and safety; types of housing; aesthetics of neighbourhoods; mixed-use neighbourhoods. Each group discussion are about 2-3 hours long and interviews are approximately 45-90 minutes long. After data collection is completed, each neighbourhood are scored according to the QASES scale by deciding on

an aggregated SES score of housing types, assets and community outlook. The SES ascription is informed by the research activities in Smithson.

Counting the hours needed for group discussions and key informant interviews amounts to 2 working days (e.g., 12.5 hours for group discussions and 4 hours for interviews). In addition, approximately one day should be spent to conduct observational activities for each neighbourhood (8 days in total). Therefore, these activities require spending 10 days in the community to collect data, together with ascribing SES to each neighbourhood. The data, experiences and reflections of the activities inform the decision-making of the SES ascription. These activities are done by 3 qualified and trained researchers (2 senior and 1 junior). The rate of wages per hour for a junior researcher is R69.71, compared to R152.74 for a senior researcher (Table 7). The senior researchers are required due to their level of training and qualifications. For instance, they should have the necessary skills to conduct and lead in-depth interviews and group-discussions; to adequately process the data (e.g., transcribing of interviews and data storage); and to perform data analysis and report writing in a timely manner. In Table 6, the total costs of labour and required time for QASES data collection are stated.

Table 6: Resource requirements (in terms of the number of staff, staff costs and time) for 2 study options

Resources	**QASES**	**Standard SES survey**
Output (number of neighbourhoods)	8 neighbourhoods	8 neighbourhoods
Staff requirements	1 team of 3 researchers (1 junior and 2 senior)	2 pairs of fieldworkers (4) 1 supervisor
Total time for research activity	10 days	12 days

Cost of staff based on qualification and working days	R5 576.90 (1 research assistant) R24 438.40 (2 social scientists)	R20 104.80 (4 fieldworkers) R21 932.40 (1 supervisor)
Total cost of employment	R30 015.30	R42 037.20

Table 7: Stellenbosch University remuneration post-level rates for required staff

Post-level bands (2020)	**QASES**		**Standard SES survey**	
Level 17 rates			Fieldworker	R108 900 p/a
				R9 075 p/m
				R2 094 p/w
				R418.85 p/d
				R52.36 p/h
Level 14 rates	Research Assistant	R145 000 p/a		
		R12 083 p/m		
		R2 788 p/w		
		R557.69 p/d		
		R69.71 p/h		
Level 10 rates	Social Scientist	R317 700 p/a		
		R26 475 p/m		
		R6 110 p/w		
		R1 221.92 p/d		
		R152.74 p/h		
Level 8 rates			Senior research supervisor	R475 200 p/a
				R39 600 p/m

			R9 138 p/w
			R1 827.70 p/d
			R228.46 p/h

3.2.1.2. Example 2: Large study sample in the city of Harare, Zimbabwe

The national Department of Health in Zimbabwe want to understand SES in relation to accessing health services across the country's capital of Harare. The estimated population size of Harare is 1 million people. There are 100 communities in Harare with each community having 16 neighbourhoods. The number of households per community were estimated at 4000; 250 per neighbourhood. The Department of Health approached SES data collection in two ways. Using option A, they implement individual data collection from a random sample of households across Harare using a standard SES survey. Or using option B, they implement QASES in Harare. I determined the labour force and costs of applying the two different methods while keeping the time as a fixed variable. For example 2, the time frame to know the distribution of SES in Harare was 6 months.

3.2.1.2.1. Option A: standard SES surveys in Harare

Approximately 20 questionnaire items are included in the individual-level SES survey. The range of questions relates to education, occupation and income (in the form of a wealth index).

When applying a standard SES survey at individual-level in the city, I estimated that surveys should be conducted with 400 randomised households out of 4000 households per community, i.e., 40 000 surveys in total. These 400 randomised households have to be visited to conduct individual surveys with. This is to ensure that when the results of the surveys are aggregated, representative results are yielded for each community. The estimation rates remain the same as the small study sample scenario: a pair of fieldworkers can complete 8-9 surveys per day assuming that recruitment, plus consent, plus survey completion would take approximately 45 minutes per participant within an 8-hour working day. To conduct 400 surveys per community, an estimated 4 pairs of fieldworkers are required to collect data within 12 days. In 6 months, these 4 pairs should have completed 10 communities. Therefore, as indicated in Table 8, to complete data collection in 100 communities over

6 months, 40 pairs of fieldworkers are required. Every 4 pairs of fieldworkers require a research supervisor (10 supervisors in total). In Table 8 the wage rates were determined for 80 fieldworkers that worked 120 days (6 months). Similarly, the wage rates were determined for 10 supervisors who worked 120 days.

It is worth mentioning that the same issues apply as in the small sample study regarding time lost to find an eligible participant at home when attempting the first visit. The only difference is that the stakes are much higher due the magnitude of the study. The process of planning should be much more rigorous to ensure that sufficient fieldworkers are appointed and that they are adequately managed. It should also be taken into account that backup fieldworkers must be appointed or available to stand in for those who are absent. Due to the vast number of workers required, it becomes more difficult to manage and oversee that the process runs efficiently.

3.2.1.2.2. Option B: QASES in Harare

The QASES research activities were adapted to the type of information required. Observations take place in each community in terms of the neighbourhood structures. This is done by observing (1) entry-exit points, (2) economic places, markets and services (e.g. gendered spaces like bars, hair salons/barbers), (3) transect walk from local clinic/health-care centre/hospital (public and private) to observe housing types, infrastructure, available health facilities and movement of people in close proximity to these different health facilities, (4) visit transport depots to observe movement (5) approach people on the streets and have short conversations with them regarding their understanding of Harare's class structures, or specifically of their communities. The topics that are covered are the same as mentioned in the small study example (see section 3.2.1.1.2).

The topic areas for group discussions and key informant interviews are focused on: what is people's understanding of SES in their communities. Again, see section 3.2.1.1.2 for the relevant topic areas pertaining to in-depth discussions with local community members and stakeholders. Group discussions take place in the streets and at the government or organisation facilities with government officials or NGO members (e.g., social development and other experts), old men, young men, old women, young women (5 group discussions). Key informant interviews with local stakeholders are

also done with experts in socio-economic profiling in the community, like government officials (e.g., social development) and members of NGO's (~4 interviews). After data collection, the QASES scale are used to give aggregated SES scores to each neighbourhood based on housing types, assets and community outlook.

To apply QASES to the 100 communities (each community having 16 neighbourhoods), the estimates are summarised in Table 8. In order to complete all the activities in 1 community in 12 days, I estimated that 1 team should be employed that consists of 4 researchers (2 junior research assistants and 2 senior social scientists). The team are split in 2 pairs (1 junior and 1 senior researcher per pair). To make sense of QASES in one community with 16 neighbourhoods, it would take 2 days to conduct group discussions and key informant interviews similarly to the small study area scenario. It is estimated that the observational activities would take 8 days in one community when 2 pairs of researchers conduct the observations. In one community, pair 1 can do half of the interviews and group discussions while pair 2 does observations. Pair 2 can then also do the remaining interviews and group discussions while pair 1 does observations simultaneously. Therefore, the research activities are divided among the 2 pairs in order to complete 1 community in 12 days, and to ensure that both pairs get exposure in all the research activities.

Afterwards, both pairs can compare their findings in order to inform one another when scoring SES for each neighbourhood. This will take an additional 2 days for the pairs to work together to ascribe SES. In total, to collect data in 12 days for 1 community, 2 pairs of 2 researchers each are required. To complete data collection in 6 months (120 working days) for all 100 communities, 2 pairs (1 team) can complete 10 communities. Therefore, 10 teams (or 20 pairs) are required to collect data. This means that there will be 20 senior researchers and 20 junior researchers. The cost of employment for the 6 months are determined as indicated in Table 8.

The data collection planning would be intensive and detailed-oriented for the 10 teams. This might require additional workers to help with planning and implementation of the activities. I however, only include the labour efforts for data collection and exclude the planning that concerns

contacting and finding participants for the group discussions and interviews. I created the scenarios assuming that planning has already been concluded when data collection started.

Table 8: Resource requirements (in terms of the number of staff and staff costs) in 6 months (120 working days) for 2 study options

Resources	**QASES**	**Standard SES survey**
Output (number of neighbourhoods)	100 communities with 16 neighbourhoods each	100 communities with 16 neighbourhoods each
Total days per community	12	12
Staff requirements per community	1 team of 2 pairs (4 researchers)	1 team of 4 pairs of fieldworkers 1 supervisor
Total staff requirements for 6 months	10 teams of 2 pairs each (40 researchers)	40 pairs of fieldworkers (80 fieldworkers) 10 supervisors
Cost of staff for 6 months	R3 177 000 (20 social scientists) R1 449 960 (20 research assistants)	R4 356 000 (80 fieldworkers) R2 367 000 (10 supervisors)
Total cost of employment	R4 626 960	R6 732 000

3.2.1.3. Summary of the hypothetical applications

I aimed to determine the labour costs and time required to complete data collection using the two different methods. In the small study sample scenario, I have indicated that it would be quicker and less expensive to conduct a SES study using QASES. Similarly, when the time frame is kept constant to collect SES data in a large study sample, it was cheaper to employ researchers for QASES. Also, less researchers were necessary to conduct the QASES study compared to the standard SES approach. Ultimately, the greatest benefit of QASES is that there is no need to retry visit attempts based on a pre-specified list of randomly sampled households since it is a once-off data collection tool. A standard SES survey tool is dependent on the number of surveys completed, even if that means the fieldworkers have to re-attempt to find the participants if they were initially unsuccessful. I must emphasise that the pre-planning and implementation of the different studies would be equally complicated and rigorous. However, the difference is that SES can be determined immediately through the QASES activities, while data extraction and analysis must still happen after survey data completion in order to determine SES for each neighbourhood. Overall, the process of conducting QASES are more efficient than the process of conducting individual SES surveys.

3.3. Accuracy of the QASES method

This section is divided in two parts. Firstly, I explore the correlations between the unchanged/unaltered QASES and PC0 SES data, and correlations where outliers have been excluded. This is to determine concurrent validity of the QASES method through the process of correlation between two different sets of scores (Loewenthal & Lewis, 2001). Secondly, I explore the level of agreement between QASES values and PC0 values by transforming the data of QASES (total) to increase agreement/uniformity to that of PC0. Additionally, I also manually changed the PC0 data from interval to ordinal, by applying the round formula on excel and adding 3 to each score. The formula is – Round((PC0+3)*2). This enabled me to compare the transformed QASES total scores to that of the transformed PC0 scores using Wilcoxon Rank Sum Test. I made a graphic representation as shown in Figure 22 of the ordinal variables – PC0 and QASES – looking at the aggregate of the scatter plot to see how many scores there are of each value (from 1-8).

3.3.1. Statistical analyses of QASES and PC0

3.3.1.1. Correlate individual QASES sub-scales

In section 2.9.5, I explained the methodology to ascribe SES in the 9 study communities using the QASES tool. As mentioned, this tool consists of three sub-scales – housing, assets, and community outlook. Researchers assigned the neighbourhood zones a score out of 0 to 4 (0 being very poor and 4 being very good) for each sub-scale. They then combined the scores of the sub-scales to aggregate the average SES per neighbourhood zone. By correlating the individual sub-scales of QASES, I wanted to get a sense of the internal coherency between the scales. This is to inspect whether each sub-scale measure what they are supposed to measure – SES. The consistency of identifying strong or weak associations between the sub-scales help to clarify further correlations between QASES and PC0 SES. Firstly, I present bar graphs of each of the QASES sub-scales that showcase the range and distributions of the data. Secondly, I perform Spearman's correlation test on the sub-scales (i.e., housing and assets, housing and community outlook, assets and community outlook).

3.3.1.1.1. Bar graphs of individual QASES sub-scales

In Figure 7 the count of QASES sub-scale scores of housing are illustrated for the 9 study communities. Each category on the x-axis is scored between 0-4 (0 being very poor and 4 being very good). The highest count/percentage values are in the range of 2 (53%) followed by 3 (45%). The distribution is slightly left skewed, meaning that across the 9 communities, most housing structures are adequate to good.

In Figure 8 the count of QASES sub-scale scores of assets is illustrated for the 9 study communities. Each category on the x-axis is scored between 0-4. The highest count/percentage values are in the range of 3 (56%) followed by 2 (53%). Notice that there are no scores for the category of 0 (very poor) because it can be assumed that all households have basic assets. Most of the values fall in the centre of the distribution with fewer scores in the extremes, making the distribution symmetric. Across the 9 communities, most household assets are adequate to good.

In Figure 9 the count of QASES sub-scale scores of community outlook are illustrated for the 9 study communities. Each category on the x-axis is scored between 0-4. The highest count/percentage values are in the range of 1 (60%) followed by 2 (58%). The values are relatively centrally distributed which implies symmetry. However, the distribution is skewed to the right. Community outlook across the 9 communities is predominantly poor to adequate.

3.3.1.1.2. Test of correlations between QASES sub-scales

In Table 9 the Spearman correlation between the QASES sub-scales of housing and assets indicates a moderately strong, positive correlation – $r = 0.617$, $n = 142$, $p = 0.000$. In Table 10 the Spearman correlation between the QASES sub-scales of housing and community outlook indicates a very strong, positive correlation of $r = 0.739$, $n = 145$, $p = 0.000$. In Table 11 the Spearman correlation coefficient between the QASES sub-scales of community outlook and assets indicates a moderately strong, positive correlation – $r = 0.485$, $n = 145$, $p = 0.000$.

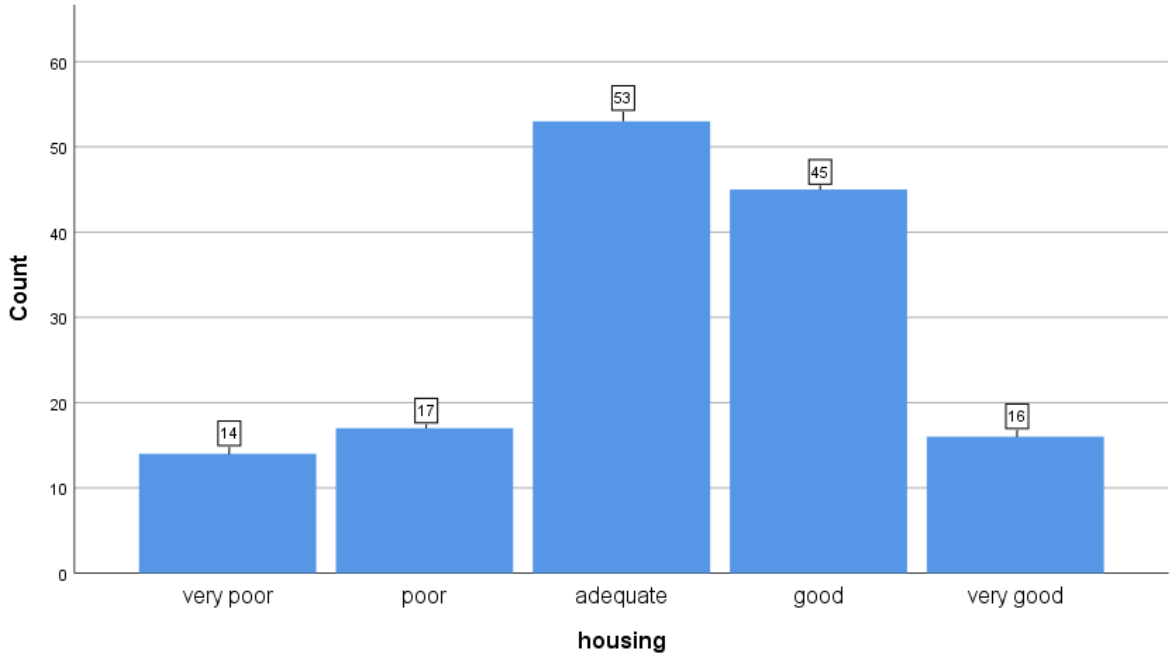


Figure 7: Bar graph of QASES sub-scale of housing

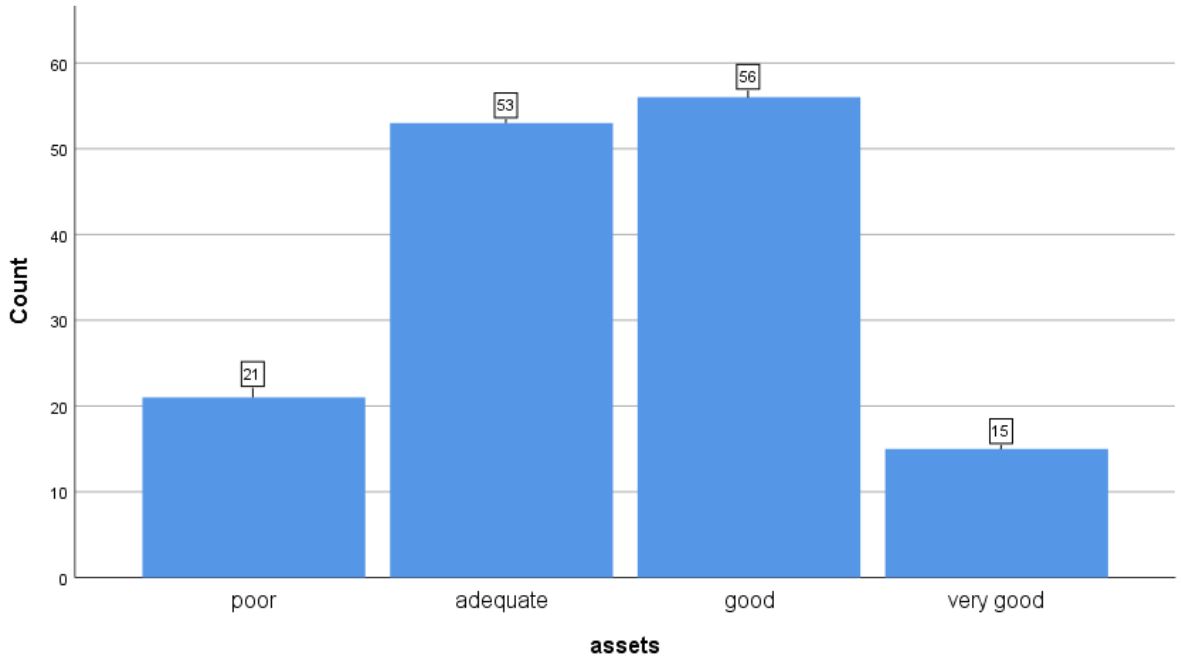


Figure 8: Bar graph of QASES sub-scale of assets

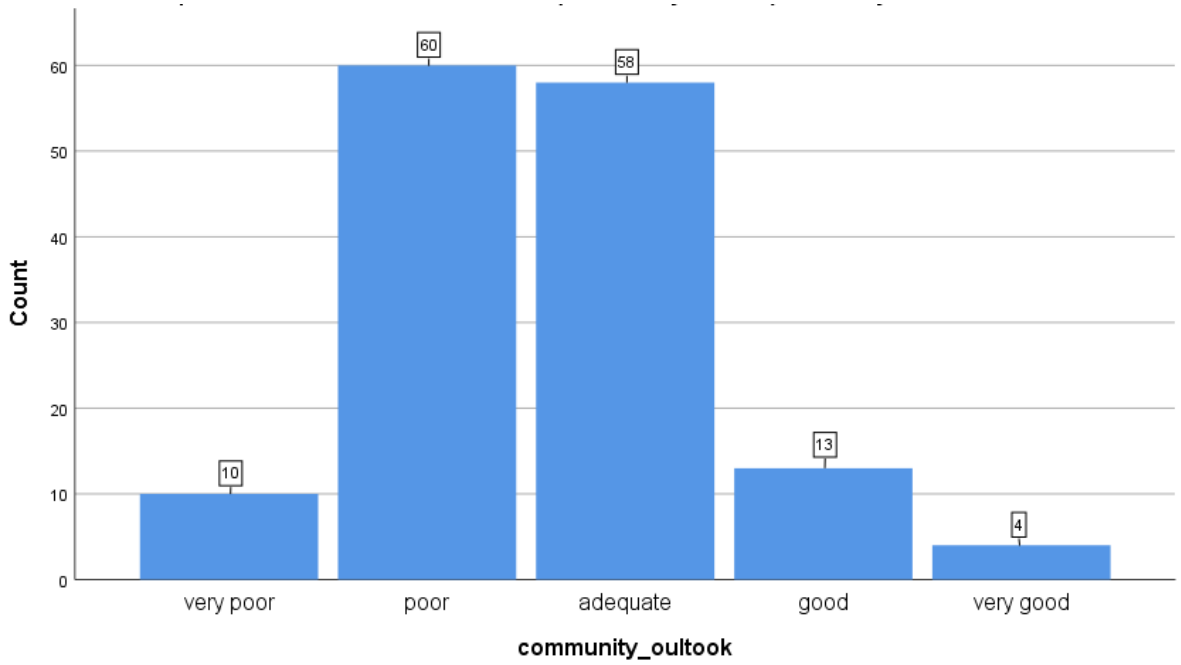


Figure 9: Bar graph of QASES sub-scale of community outlook

Table 9: Spearman correlation of QASES housing and QASES assets

Correlations

			QASES_housing	QASES_assets
Spearman's rho	QASES_housing	Correlation Coefficient	1.000	.617**
		Sig. (2-tailed)	.	.000
		N	145	145
	QASES_assets	Correlation Coefficient	.617**	1.000
		Sig. (2-tailed)	.000	.
		N	145	145

**. Correlation is significant at the 0.01 level (2-tailed).

Table 10: Spearman correlation of QASES housing by community outlook

Correlations

			QASES_housing	QASES_community_oultook
Spearman's rho	QASES_housing	Correlation Coefficient	1.000	.739**
		Sig. (2-tailed)	.	.000
		N	145	145
	QASES_community_oultook	Correlation Coefficient	.739**	1.000
		Sig. (2-tailed)	.000	.
		N	145	145

**. Correlation is significant at the 0.01 level (2-tailed).

Table 11: Spearman correlation of QASES community outlook and assets

Correlations

			QASES_community_oultook	QASES_assets
Spearman's rho	QASES_community_oultook	Correlation Coefficient	1.000	.485**
		Sig. (2-tailed)	.	.000
		N	145	145
	QASES_assets	Correlation Coefficient	.485**	1.000
		Sig. (2-tailed)	.000	.
		N	145	145

**. Correlation is significant at the 0.01 level (2-tailed).

3.3.1.2. Total QASES and PC0

I illustrate the correlations between the total QASES scores (determined as the sum of the sub-scale scores) and the PC0 scores. Firstly, I present bar graphs of QASES and PC0 SES in order to illustrate the distributions and ranges of the data. Then I present scatter plots of the two datasets to determine the relationship in terms of the shape and direction of the data. This is important to demonstrate by means of the data points whether the scores align. The scatter plot includes all 145 data points. Lastly, I performed Spearman's correlation coefficient to determine the strength of association between QASES and PC0 SES. I performed two tests – one with all data points, and the other with the three outliers removed. I sought to explore how much the outliers affected the association between the scores.

3.3.1.2.1. Bar graphs of total QASES and PC0 scores

In Figure 10 the total QASES scores of housing, assets and community outlook are illustrated for the 9 study communities, meaning the sum of all sub-scale scores per community zone. The highest percentage values fall within the range of 6-7 (31% and 34%). Therefore, most of the values fall in the centre of the distribution with fewer scores in the extremes, making the distribution symmetrical. Across the 9 communities, most households have a poor to adequate SES.

Figure 11 shows the frequencies in percentages for the PC0 SES scores. From the graph, most of the scores gather within -1 to 1, with the mean being 0.07. The mean indicates that there is less dispersion in the data and therefore the data have a central tendency. The distribution is skew towards the left with a long tail to the left. Some of the skewness is attributable to the outliers in the data.

3.3.1.2.2. Scatter plot and correlation test between total QASES and PC0

Figure 12 indicates a strong, positive, linear relationship/association between X (total QASES) and Y (PC0 SES). The plot presents few outliers to the left. Outliers are expected because there exists variability to how the data for each study were collected. As also explained in the methods section, these outliers might be caused by the process of combining and altering the BBS and PC0 zones to correct the alignment of the zones (see sections 2.11.1.1 and 2.11.1.3). In Appendix J, I

excluded the outliers in the scatter plot. As indicated, the scatter plot looks similar when outliers are removed.

In Table 12 the Spearman correlation coefficient was computed for the total QASES and PC0 SES. There is a strong, positive correlation between the two variables, r = 0.762, n = 145, p = 0.000. In Table 13 Spearman correlation coefficient was computed to assess the relationship between the total QASES scale and PC0 SES (excluding outliers). There is a strong, positive correlation between the two variables, r = 0.753, n = 142, p = 0.000.

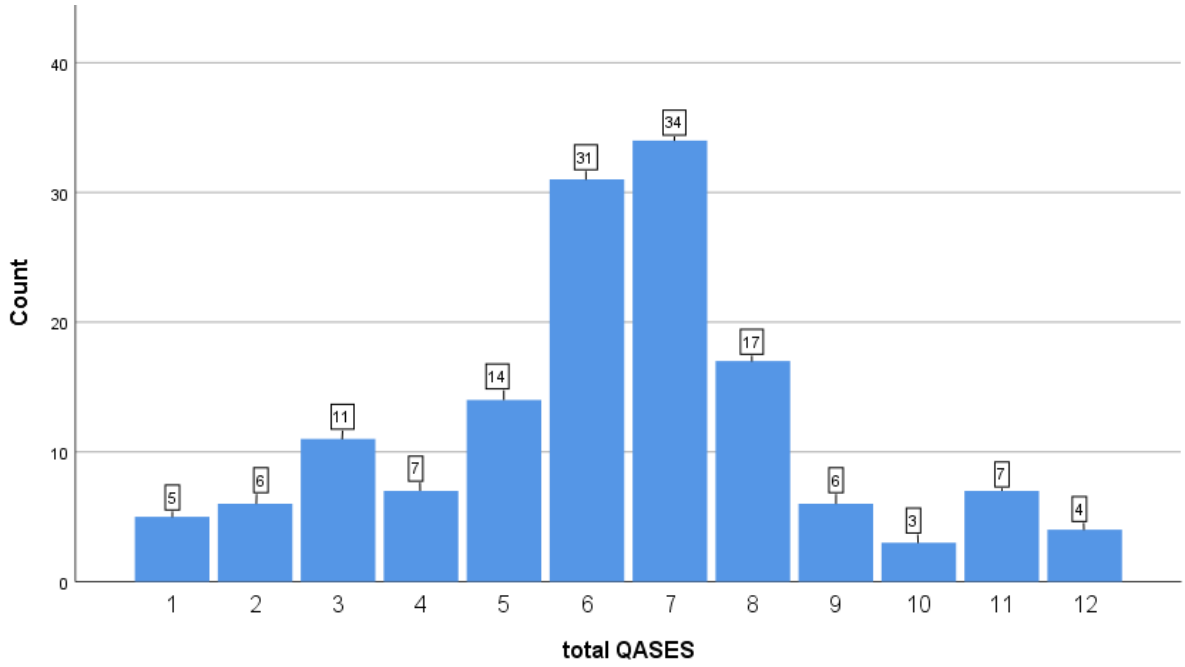


Figure 10: Bar graph of total QASES

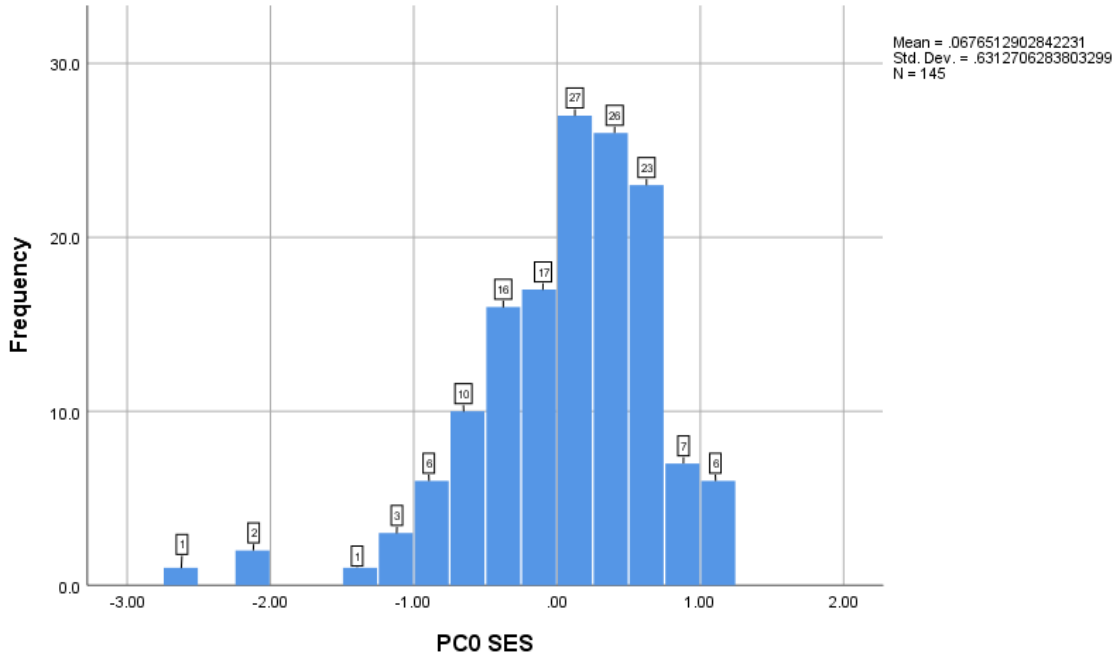


Figure 11: Bar graph of PC0 SES

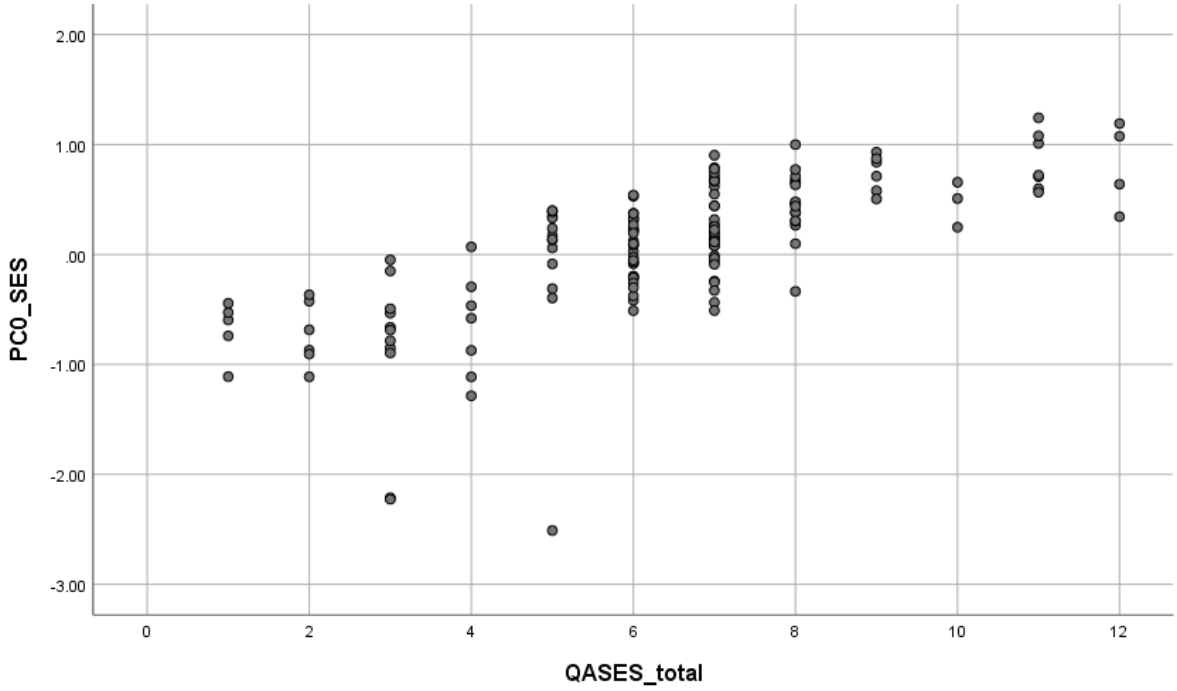


Figure 12: Simple scatter of total QASES by PC0 SES (including outliers)

Table 12: Spearman correlation of QASES total and PC0 SES

Correlations

			QASES_total	PC0_SES
Spearman's rho	QASES_total	Correlation Coefficient	1.000	.762**
		Sig. (2-tailed)	.	.000
		N	145	145
	PC0_SES	Correlation Coefficient	.762**	1.000
		Sig. (2-tailed)	.000	.
		N	145	145

**. Correlation is significant at the 0.01 level (2-tailed).

Table 13: Spearman correlation of PC0 SES and total QASES (outliers removed)

Correlations

			QASES_total	PC0_SES
Spearman's rho	QASES_total	Correlation Coefficient	1.000	.753**
		Sig. (2-tailed)	.	.000
		N	142	142
	PC0_SES	Correlation Coefficient	.753**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

3.3.1.3. Individual QASES sub-scales to PC0 total

I illustrate the correlations between each individual QASES sub-scale and PC0 SES. Firstly, I do this by creating Box-and-Whisker plots of the QASES sub-scales compared to PC0. The use of box plots contributes to assessing the symmetry or skewness and dispersion of the data, especially for drawing comparisons of these features in the two datasets. Furthermore, I run the correlation test of Spearman between each individual QASES sub-scale and PC0. As similarly done by comparing each QASES sub-scale to one another, these tests of associations would highlight trends in terms of the coherency in the QASES scale.

3.3.1.3.1. Box plot and correlation test of QASES housing versus PC0

For the first housing category – 0, very poor – the median is -0.67 (Figure 13). The variability of the data around the median indicates a tight distribution (short box). The median is in the centre of the box, indicating that the data are evenly distributed on either side of the median, making the distribution symmetric. The data are slightly skewed to the lower half of the box. Therefore, the length of the box plot is longer to the lower half, meaning that more data condensed on the upper end (25% and up) of the PC0 SES scale, with values spread out on the bottom 25%.

When looking at the total distribution of the data, as QASES increases, PC0 SES also increases. For instance, the accumulation of values that are scored ‘adequate’ (2) for QASES, are similar on the PC0 SES scale (in the centre at .00). For each QASES category the variability differs, since most of the categories have a greater variability (longer boxes), than category 0 on QASES. Two case numbers are identified as outliers in category 2 (adequate) and category 4 (very good). These values are considered as outliers because they deviate from the remaining observations. Specifically, lower scores were ascribed on the PC0 scale compared to the QASES housing sub-scale. The skewness of each data category is generally directed to the lower end. In conclusion, the longer boxes (high variability) and lower skewedness of the data indicate that the values for QASES housing sub-scale are underestimated compared to PC0 SES.

In Table 14 Spearman's correlation coefficient was computed to assess the relationship between the housing QASES sub-scale and the PC0 SES tool (excluding the outliers). There is a strong, positive correlation between the two variables, r = 0.768, n = 142, p = 0.000.

3.3.1.3.2. Box plot and correlation test of QASES assets versus PC0

For the first assets category – 1, poor – the median is -0.69 (Figure 14). The variability of the data around the median indicates a tight distribution (short box). The median is in the middle of the box, indicating that the data are evenly distributed on either side of the median, making the distribution symmetric. The data are slightly skewed to the upper half of the box (75%). The inter-quartile range (IQR) (minimum, Q1, median, Q3, maximum) are centrally distributed.

When looking at the total distribution of the data, as QASES increases, PC0 SES also increases. Except, for categories 2 and 3 (adequate and good), there is only a slight increase between the two categories. The extreme outlier skews the bottom 25% of category 2 lower. Most of the categories have a greater variability (longer boxes), except category 4 (very good). Even though category 4 data are tightly distributed, the median is located more towards the bottom of the box, making the distribution for the category asymmetrical.

In Table 15 Spearman's correlation coefficient was computed to assess the relationship between the assets QASES sub-scale and the PC0 SES tool (excluding outliers). There is a moderately strong, positive correlation between the two variables – r = 0.540, n = 142, p = 0.000.

3.3.1.3.3. Box plot and correlation test of QASES community outlook versus PC0

For the first community outlook category – 1, very poor – the median is -0.51 (Figure 15). The variability of the data around the median indicates a tight distribution (short box). The median is located closer to the lower end of the box towards Q3 meaning that the data are not evenly distributed on either side of the median, or asymmetric. This category has 2 outliers on both ends of the maximum and minimum values. The data are skewed on the upper half of the box (75%), but more condensed from the median downwards (median, Q3 and maximum).

For the entire data representation and distribution, when QASES increases, PC0 SES also increases. Most of the categories have a greater variability (longer boxes and whiskers), except category 1 (very poor). Even though category 1 data are tightly distributed, the median is located more towards the bottom of the box (close to Q3), making the distribution for the category asymmetrical.

In Table 16 Spearman correlation coefficient was computed to assess the relationship between the QASES community outlook sub-scale and PC0 SES. There is a strong, positive correlation between the two variables – r = 0.624, n = 142, p = 0.000.

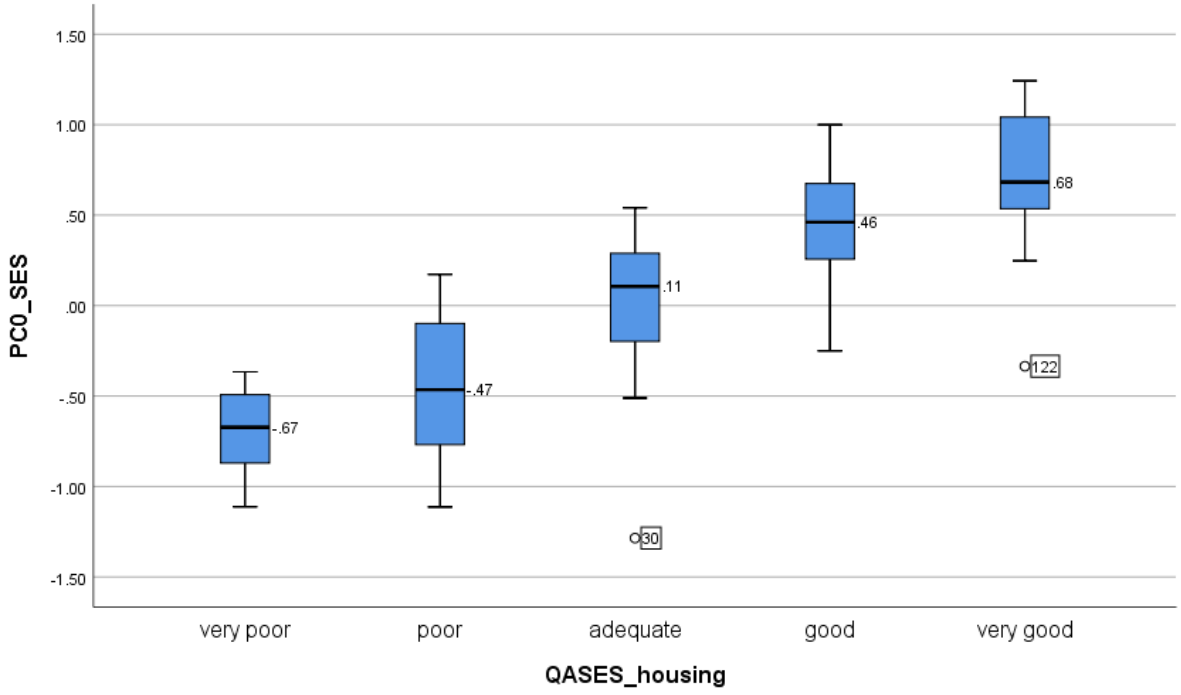


Figure 13: Box plot of the QASES housing category (0, very poor to 4, very good) by PC0 SES

Table 14: Spearman correlation of QASES housing and PC0 SES (excluding outliers)

Correlations

			QASES_housing	PC0_SES
Spearman's rho	QASES_housing	Correlation Coefficient	1.000	.768**
		Sig. (2-tailed)	.	.000
		N	142	142
	PC0_SES	Correlation Coefficient	.768**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

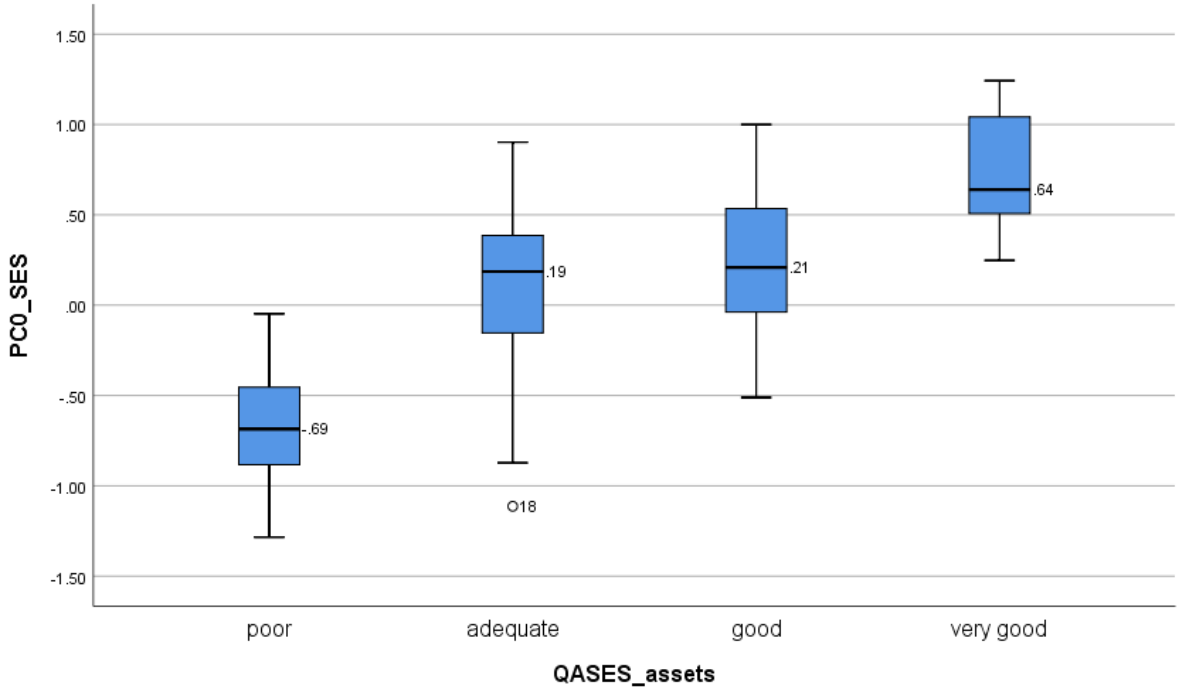


Figure 14: Box plot of QASES sub-scale (assets) by PC0 SES

Table 15: Spearman correlation of QASES assets and PC0 SES (excluding outliers)

Correlations

			PC0_SES	QASES_assets
Spearman's rho	PC0_SES	Correlation Coefficient	1.000	.540**
		Sig. (2-tailed)	.	.000
		N	142	142
	QASES_assets	Correlation Coefficient	.540**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

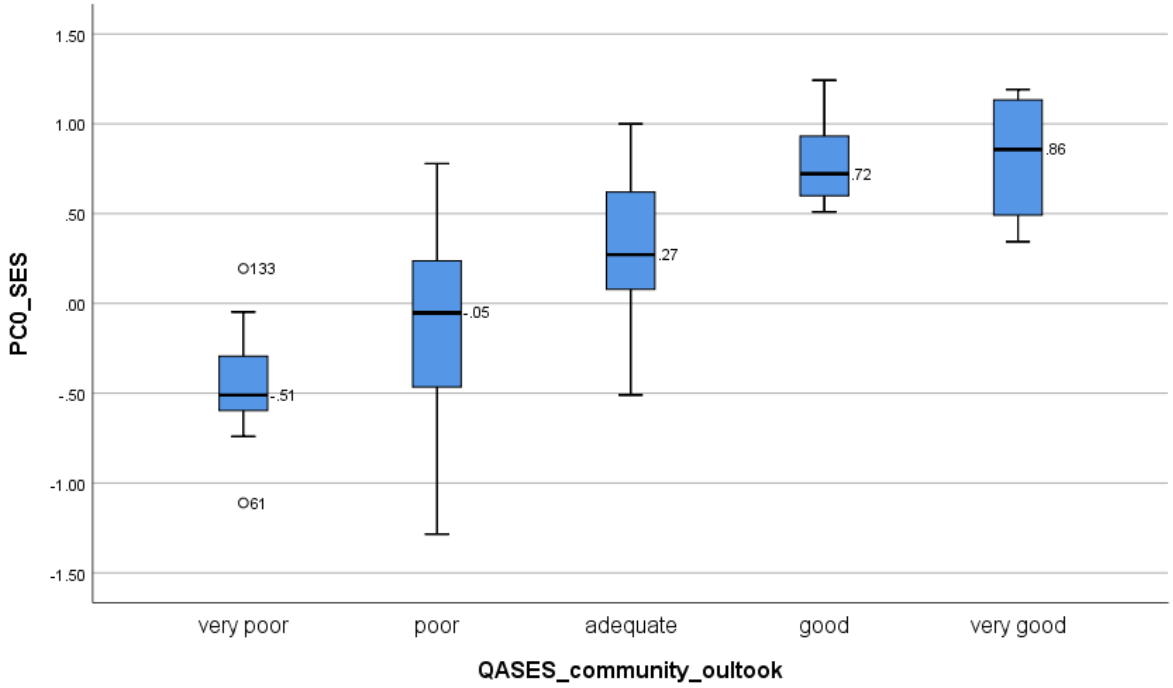


Figure 15: Box plot of QASES sub-scale (community outlook) by PC0 SES

Table 16: Spearman correlation of QASES community outlook and PC0 SES (excluding outliers)

Correlations

			PC0_SES	QASES_community_oultook
Spearman's rho	PC0_SES	Correlation Coefficient	1.000	.624**
		Sig. (2-tailed)	.	.000
		N	142	142
	QASES_community_oultook	Correlation Coefficient	.624**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

3.3.1.4. Combined QASES sub-scales to PC0 total

I combined the QASES sub-scales (e.g., assets plus housing) and performed correlation analysis. The aim of doing this was, again, to justify which QASES sub-scales are more coherently similar when compared to the PC0 data. For each combined sub-scales, I firstly made bar graphs to illustrate the distribution and range of the data. This is followed by scatter plots of the combined QASES sub-scales data that are compared to the PC0 data. I did this to illustrate the direction, shape and distribution of the correlated data. Finally, for each combined QASES sub-scale data versus PC0 data, I calculated the strength of the association using Spearman's correlation coefficient.

3.3.1.4.1. Bar graph, scatter plot and test of correlation for QASES combined sub-scales (*housing plus assets*) and PC0)

Figure 16 shows the frequencies in percentages for the QASES combined sub-scales of housing plus assets. From the graph, the mean of the distribution is 4.67, which is the average score on the scale of 0 to 8. The range of the data are central, with most scores falling in the categories of 4, 5 and 6. Overall the data are normally distributed.

Figure 17 shows a strong, positive, linear association of the combined QASES sub-scales (housing plus assets) to PC0 SES. There are some visible outliers on the QASES categories of 3 and 7, where a higher score was given (in combination) for QASES than PC0 for those particular zones.

In Table 17 Spearman correlation coefficient was computed to assess the relationship between the combined QASES sub-scales of housing plus assets and PC0 SES (outliers excluded). There is a strong, positive correlation between the two variables – $r = 0.734$, $n = 142$, $p = 0.000$.

3.3.1.4.2. Bar graph, scatter plot and test of correlation for QASES combined sub-scales (*housing plus community outlook*) and PC0)

Figure 18 shows the frequencies in percentages for the QASES combined sub-scales of housing plus community outlook. The mean or average of the distribution is 3.81. The range of the

data is central, with most scores falling in the categories of 3 and 5. Overall the data are normally distributed.

In Figure 19, the association of the combined QASES (housing plus community outlook) compared to PC0 is strong, positive and linear. The most extreme outliers are found in the combined QASES categories of 3 and 8. Again, QASES (combined housing and community outlook) presented a higher score compared to PC0. The relationship of PC0 SES and the combined housing plus community outlook sub-scales shows the strongest linear relationship than the other combined QASES sub-scales.

In Table 18 Spearman correlation coefficient was computed to assess the relationship between the combined QASES sub-scales of housing plus community outlook and PC0 SES (outliers excluded). There is a strong, positive correlation between the two variables – r = 0.756, n = 142, p = 0.000.

3.3.1.4.3. Bar graph, scatter plot and test of correlation for QASES combined sub-scales (*assets plus community outlook*) and PC0)

Figure 20 shows the frequencies in percentages for the QASES combined sub-scales of assets plus community outlook. The mean or average of the distribution is 4.04. The range of the data is central and slightly right-skewed (positive). Most scores fall in the categories of 3, 4 and 5. The data are normally distributed.

In Figure 21, the association between combined QASES sub-scales (assets plus community outlook) and PC0 SES is moderately strong, positive, and linear. The visible outliers on the combined QASES categories of assets plus housing, are 1, 5, and 8. Out of all three combined sub-scale associations to PC0 SES, this is the weakest association. The common denominator thus far is the QASES sub-scale of assets that associates the weakest out of all the shown scatter plots, box plots and Spearman correlations.

In Table 19 Spearman correlation coefficient was computed to assess the relationship between the combined QASES sub-scales of assets plus community outlook and PC0 SES (outliers excluded). There is a strong, positive correlation between the two variables – r = 0.662, n = 142, p = 0.000.

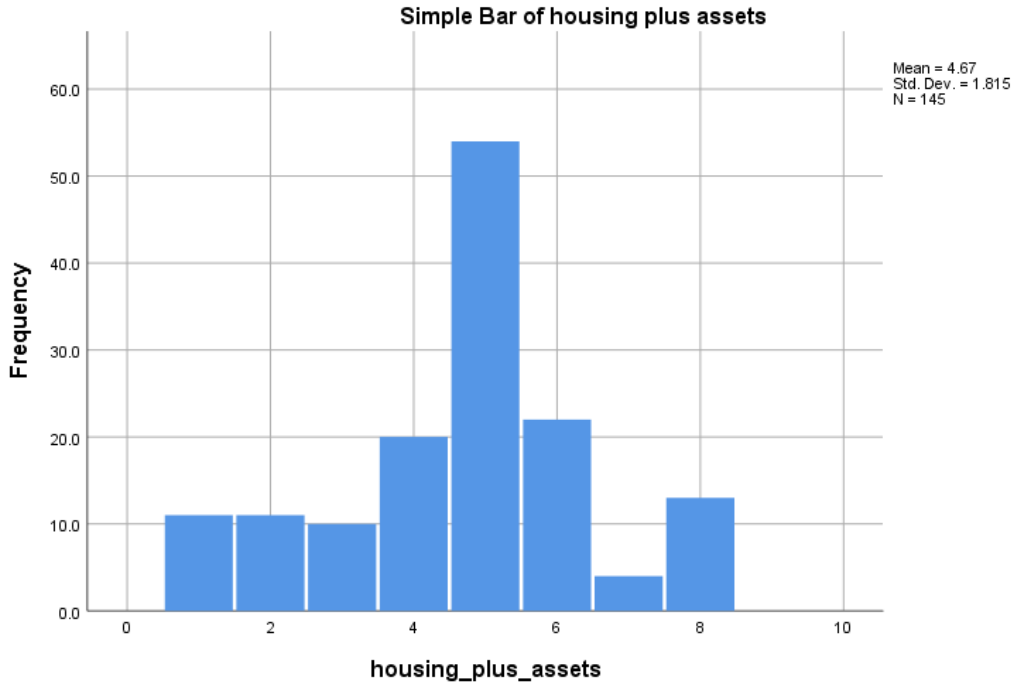


Figure 16: Simple bar graph of combined QASES sub-scales (housing plus assets)

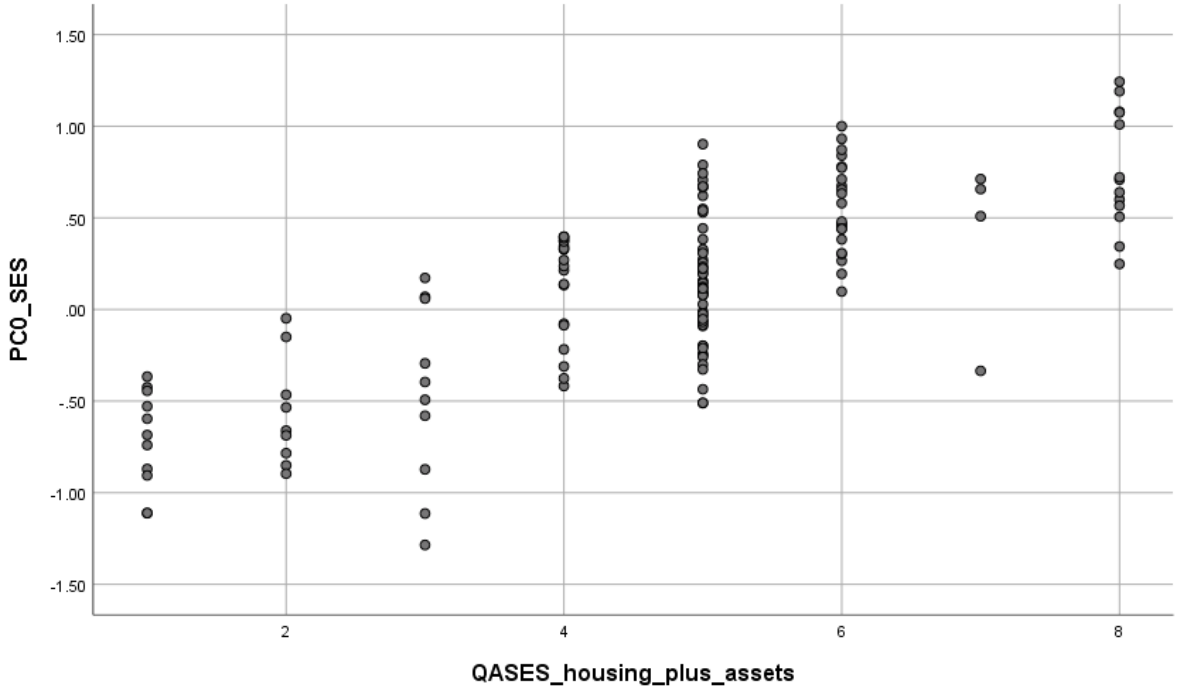


Figure 17: Simple scatter plot of combined QASES sub-scales (housing plus assets) and PC0 SES

Table 17: Spearman correlation of combined QASES sub-scales (housing plus assets) and PC0 SES

Correlations

			PC0_SES	QASES_housing_plus_assets
Spearman's rho	PC0_SES_wlthindex	Correlation Coefficient	1.000	.734**
		Sig. (2-tailed)	.	.000
		N	142	142
	QASES_housing_plus_assets	Correlation Coefficient	.734**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

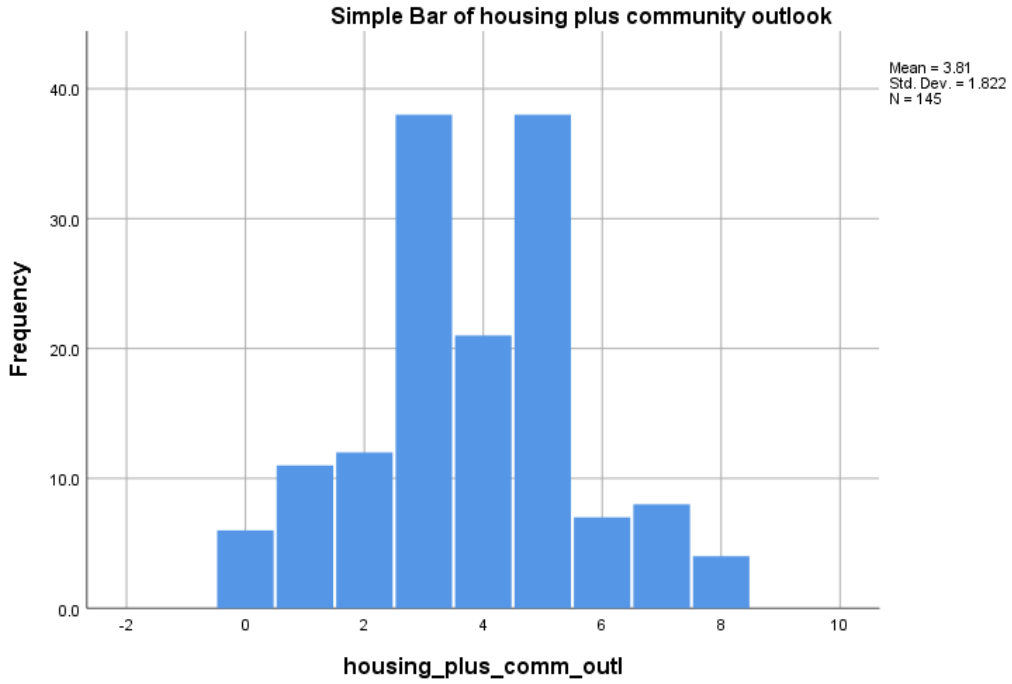


Figure 18: *Simple bar graph of combined QASES sub-scales (housing plus community outlook)*

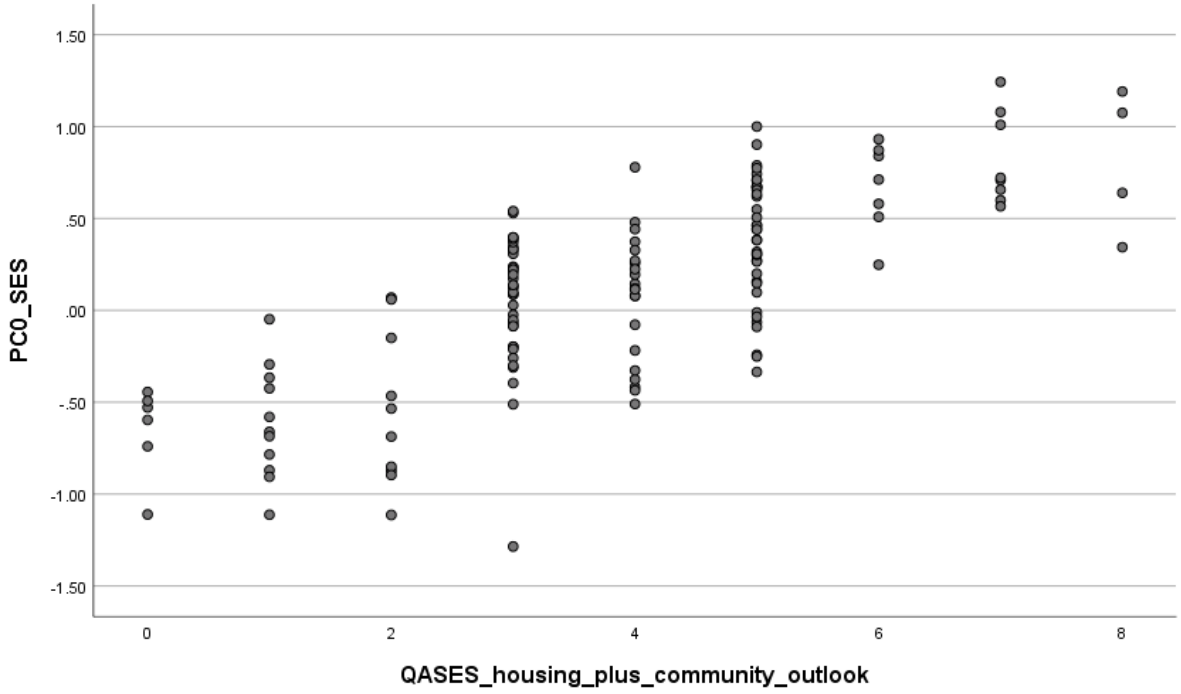


Figure 19: Simple scatter plot of combined QASES sub-scales (housing plus community outlook) and PC0 SES

Table 18: Spearman correlation of combined QASES sub-scales (housing plus community outlook)

Correlations

			PC0_SES	QASES_housing_plus_community_outlook
Spearman's rho	PC0_SES	Correlation Coefficient	1.000	.756**
		Sig. (2-tailed)	.	.000
		N	142	142
	QASES_housing_plus_community_outlook	Correlation Coefficient	.756**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

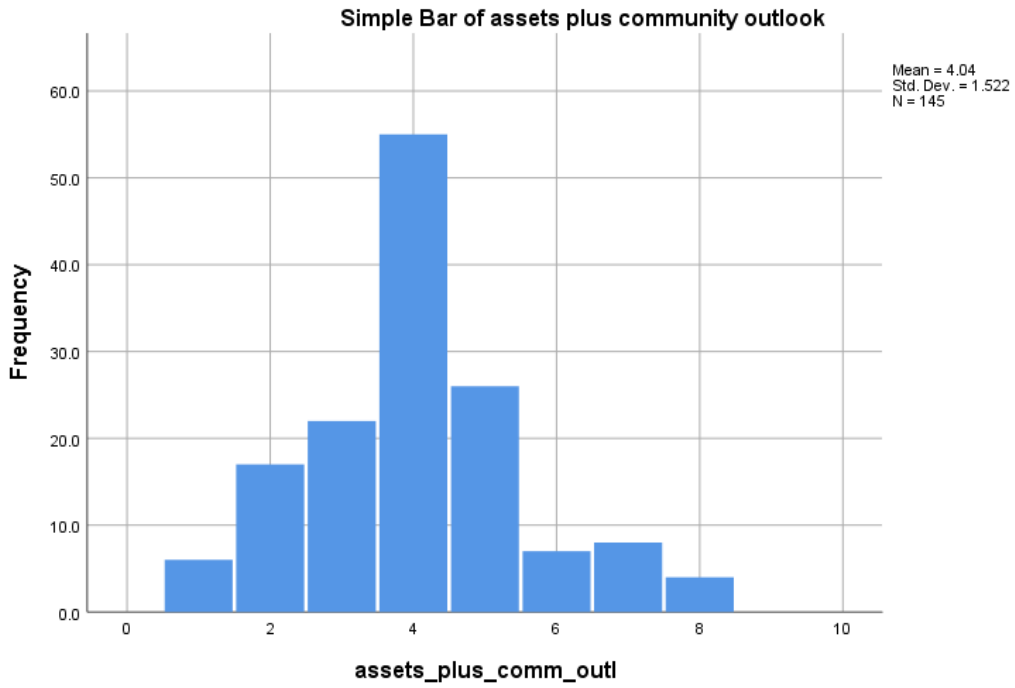


Figure 20: *Simple bar graph of combined QASES sub-scales (assets plus community outlook)*

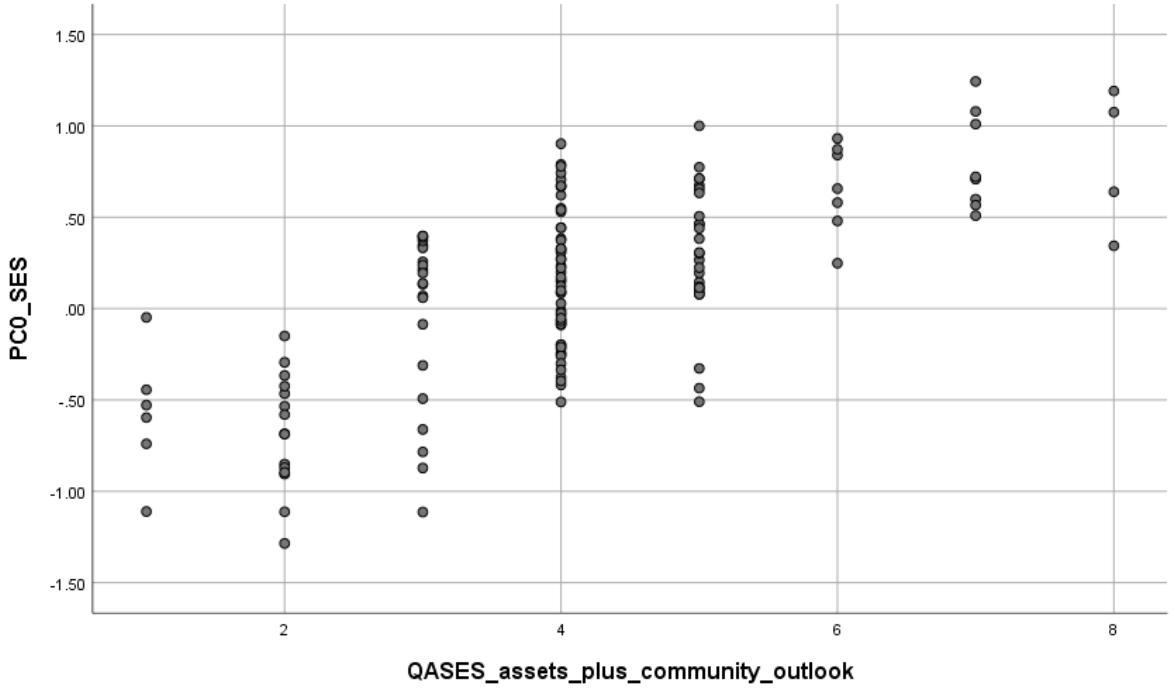


Figure 21: Simple scatter plot of combined QASES sub-scales (assets plus community outlook) and PC0 SES

Table 19: Spearman correlation of combined QASES sub-scales (assets plus community outlook) and PC0 SES

Correlations

			PC0_SES	QASES_assets_plus_community_outlook
Spearman's rho	PC0_SES	Correlation Coefficient	1.000	.662**
		Sig. (2-tailed)	.	.000
		N	142	142
	QASES_assets_plus_community_outlook	Correlation Coefficient	.662**	1.000
		Sig. (2-tailed)	.000	.
		N	142	142

**. Correlation is significant at the 0.01 level (2-tailed).

3.3.1.5. Summary of the correlations

Overall, the correlation tests for each of the different QASES variables have indicated strong, significant associations. By removing the outliers in the PC0 data, the strength of the associations was not significantly impacted. Testing the internal correlations among the individual QASES sub-scales, the assets sub-scale scored consistently lower. QASES sub-scale of assets also showed the weakest association when compared to PC0. This is followed by community outlook and then housing, which had the highest association with PC0. When the QASES sub-scales were combined, assets plus community outlook had the weakest association, whereas housing plus community outlook had the strongest association to PC0. By applying the different forms of using QASES (individual, combined and total) I aimed to highlight what may cause the strength of correlations to fluctuate. Ultimately, from each group of significance tests, the QASES method have shown to be valid (as indicated by the strong associations to the PC0 data). QASES is a representative measure of SES which it was intended to measure.

3.3.2. Statistical analysis of transformed QASES total variables and PC0 SES variables

3.3.2.1. Comparing total QASES transformed to PC0 SES transformed

The Wilcoxon rank sum test was done on the transformed data (QASES and PC0 SES) (see Appendix H for how data was transformed). The results are shown in Table 20, where the negative ranks indicate that QASES scored lower than PC0 SES 37 times. The positive ranks indicate that QASES scored higher than PC0 SES 62 times. The ties indicate that QASES and PC0 SES scored the same 46 times. The difference between the two scores was not statistically significant p=0.031 (with a 99% confidence interval). The null hypothesis is retained because there is no systematic difference in the median scores for QASES and PC0. The reason for including this test was to emphasise that there is a high rate of similarity within the scoring for both methods.

Table 20: Wilcoxon Signed Ranks Test

Ranks

		N	Mean Rank	Sum of Ranks
tot_QASES_trans - PC0SES_Round	Negative Ranks	37[a]	47.68	1764.00
	Positive Ranks	59[b]	49.02	2892.00
	Ties	46[c]		
	Total	142		

a. tot_QASES_trans < PC0SES_Round

b. tot_QASES_trans > PC0SES_Round

c. tot_QASES_trans = PC0SES_Round

Test Statistics[a]

	tot_QASES_trans - PC0SES_Round
Z	-2.158[b]
Asymp. Sig. (2-tailed)	.031

a. Wilcoxon Signed Ranks Test

b. Based on negative ranks.

3.3.2.2. Aggregate of the scatter plot (QASES total and PC0 SES) – graphical representation of ordinal data

The purpose of creating this graph in Figure 22 was to get an idea of how well the data values fit within the same categories (from 1-8) when the data was transformed to ordinal values for both QASES total and PC0 SES. The darker the blocks are in red, the more amount of values are similarly rated by each method. For instance, the highest amount in the graph is 18, which means that for both the QASES and PC0 scales, the values of 6 by 6 were scored 18 times, followed by 7 (scored 12 times). This graph gives a sense of how close the QASES scoring comes to PC0 SES. To illustrate further, the number of similar scores for both QASES and PC0 is 46 out of 145; the number of scores that were off by one value (e.g., a score of 5 for QASES and 6 for PC0) is equal to 67 out of 145. Ultimately, the success rate is 113 times out of 145 that the two scales score similar or close to similar values.

- Sum of similar scores [(4:4), (5:5), (6:6), (7:7), (8:8)]
    - 6+5+18+12+5 = 46
- Sum of similar scores – off by one [(4:3), (4:5), (5:4), (5:6), (6:5), (6:7), (7:6), (7:8), (8:7), (8:9)]
    - 1+2+7+5+7+15+8+16+3+3 = 67
- Aggregate
    - 46+67 = 113
    - Therefore, out of 145 data points, 113 times QASES scores very similar to PC0 SES

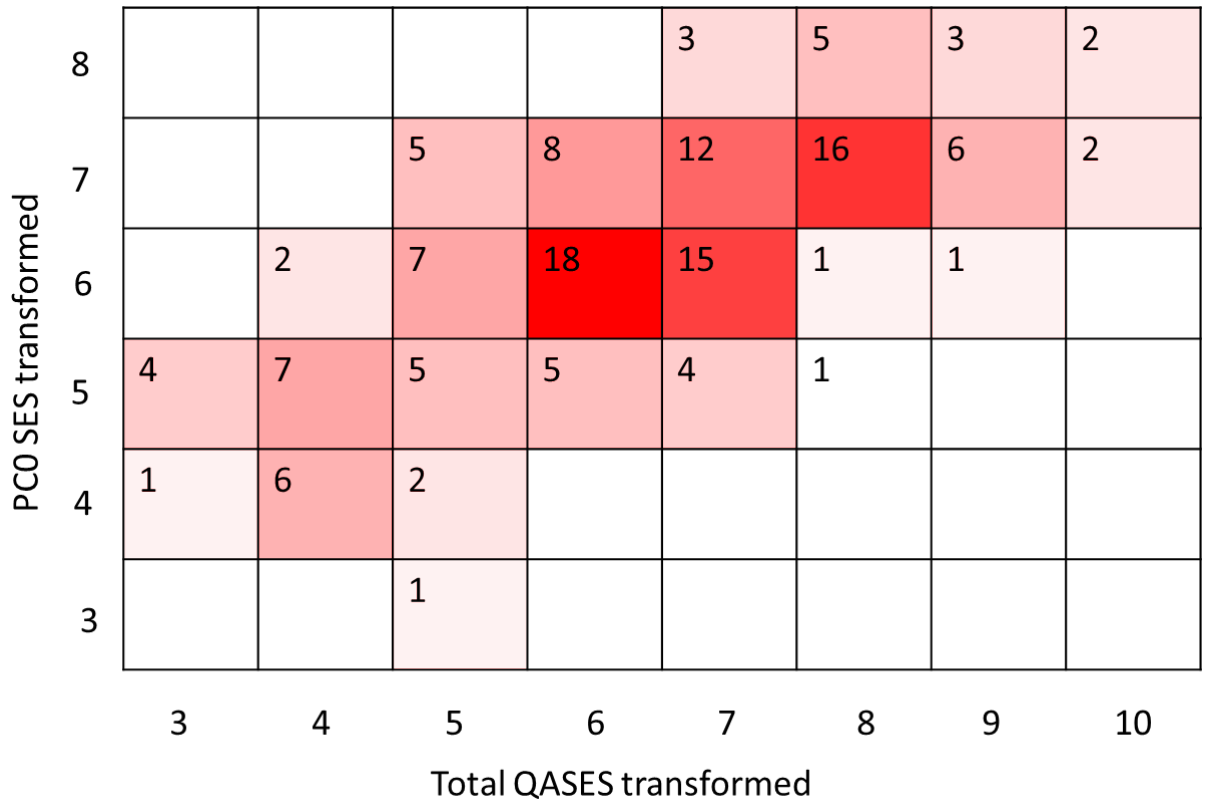


Figure 22: Aggregate of scatter plot that shows the amount of values for each variable that correlate

3.4. Findings conclusion

The analyses of comparing the QASES method to a standard SES measure (PC0 SES) have indicated that QASES is an efficient and accurate alternative method to measure SES. Through the hypothetical scenarios of comparing the QASES data collection process to a standard SES data collection process, QASES has been found to be more resource and time efficient in its application. Therefore, I reject the null hypotheses in 2.4.1, which is that there are no significant differences between QASES and an individual-level SES survey regarding costs, labour and time requirements for data collection.

For the accuracy analyses I have indicated that QASES correlates significantly to PC0. I first conducted correlations between the individual QASES sub-scales to determine the scale's internal coherency. For the correlations between housing and assets, housing and community outlook, and assets and community outlook there are sufficient evidence to conclude that significant linear relationships exist between these correlations (see Table 9, Table 10 and Table 11) because the

correlation coefficient is significantly different from zero. Therefore, the null hypotheses are rejected as stated in section 2.4.2.

Furthermore, I have found that there is a significant linear relationship between the total QASES (from all three sub-scales) and the PC0 scores (see Figure 12, Table 12 and Table 13). This concludes to significant evidence to reject the null hypothesis in section 2.4.4. Through correlating the scores of each individual QASES sub-scale to the PC0 scores, I have found sufficient evidence to conclude that significant linear relationships exist between these variables (see Table 14, Table 15 and Table 16) to reject the null hypotheses in section 2.4.5. Lastly, by combining the QASES sub-scales and correlating those scores to the PC0 total score, I have gathered sufficient evidence to conclude that significant linear relationships exist between these variables (see Table 17, Table 18, Table 19, Figure 17, Figure 19 and Figure 21). Hence, I reject the null hypotheses in 2.4.6. Overall, the null hypothesis in section 2.4.3 is rejected because there are significant statistical associations between the QASES and PC0 SES scales.

The only hypothesis that I failed to reject was the hypothesis linked to the Wilcoxon Rank Sum Test (section 2.4.7). This is due to both the transformed QASES total and PC0 scores having a median distribution not significantly different to zero (see Table 20). The transformed scores were intended to show how many times the two scales scored similarly instead of significantly differently (see Figure 22).

Chapter 4: Discussion and conclusion

4.1. A brief summary of the findings

From the literature, I identified the practice of SES scale construction and data collection as a research problem, especially in resource-scarce countries or LMICs. The standard way of collecting SES data is usually labour intensive, time consuming and expensive (Munyoro, 2018). Therefore, I explored an alternative way of measuring SES that is quicker and more operationally useful – termed QASES. The methods I used to investigate the validity of QASES, was to do determine the efficiency of applying the QASES method compared to a standard SES survey to hypothetical study areas. I measured efficiency in terms of resource requirements (labour force, costs and time) for both measures. In addition, I conducted correlation analysis (Spearman) between data collected using a standard SES measure (PC0) versus rapidly collected data (QASES). Correlation analysis allowed me to explore how close the QASES method gets to achieve similar results to PC0 SES.

Overall, the findings of the study suggest that QASES is a viable, sufficiently accurate, far less resource intensive method for determining neighbourhood level SES. There exist stipulations to this conclusion: (1) this is only true if a study team are satisfied to apply aggregate data to neighbourhoods even though their sizes differ and variability exist within neighbourhoods, (2) it is clear that internal coherency between the QASES sub-scales did not produce equally strong correlations (3) it is understood that not all the QASES variables or sub-scales deliver equally strong correlations to the standard SES measure. I have gained these perspectives from the illustrated scatter plots, box-and-whisker plots, Spearman correlations, and Wilcoxon Rank Sum Test.

The findings from the efficiency analysis was a way to explore the application of the different method types – QASES and a standard SES survey. It challenged me to think about how to apply and adapt each method to different contexts. I explored the labour intensity costs and time of using both methods to a small and large study sample. I determined that using QASES to collect data would significantly be less labour intensive, more affordable to apply and would save time compared to a standard SES-survey approach. The null hypotheses are rejected stating that there are no significant

differences in the efficiency of QASES compared to a standard SES measure (in terms of labour requirements, costs, and time) (section 2.4.1).

Through my investigation of the two different methods, I could also determine a benefit of using QASES above a standard SES survey approach: the QASES tool is not exclusively dependent on individuals as participants which ease the process of assigning an SES score based on observations. This is in contrast to PC0 or other standard SES surveys that depend on exact numbers of surveys to be completed either through an electronic survey or by paper surveys. Either way, it takes longer to extract and analyse the SES data from the standard SES survey approach than the QASES approach. This benefit refers to the ease of using QASES due to it being operationally less resource intensive.

For each set of the accuracy findings, the associations were statistically significant. Therefore, the null hypotheses are rejected that there are no association between QASES (total, individual sub-scales, and combined sub-scales) and PC0 SES (sections 2.4.2 - 2.4.6). Additionally, the Wilcoxon Rank Sum Test has confirmed that there is not a significant difference between the two measures therefore retaining the null hypothesis (section 2.4.7). The use of Wilcoxon was to illustrate the amount of times that QASES and PC0 SES have tied scores. Furthermore, the aggregate graph visually showcases the concordance of similar scores between the two measures. I wanted to demonstrate how close QASES gets to scoring similar to PC0 SES (113 out of 145 times).

4.2. The relevance of findings on how SES has been measured in health research

I aimed to address the research gap of exploring an unconventional way of measuring SES in LMICs, specifically where there exists a need to redefine SES variables and its measurement (Milenkovic *et al.*, 2014). Standardised SES measurement, in the form of lengthy surveys, is a time consuming and costly process within contexts of low resources (Filmer & Scott, 2012; Vyas & Kumaranayake, 2006). The QASES measure entails variables or sub-scales that encompass an asset-based approach to measuring SES, where information is collected on durable assets ownership (e.g., stove, refrigerator, TV, etc.) in conjunction with housing characteristics (e.g., structure material, size,

floor and roof materials, etc.) (Howe *et al.*, 2008). This combined nature of assets, housing characteristics and amenities are also termed a wealth index (Howe *et al.*, 2012; Kabudula, Houle, Collinson, Kahn, Tollman & Clark, 2017).

The only different sub-scale is community outlook which is used to monitor the well-being of LMICs (Divingri & St. Clair, 2017). A community outlook survey has been developed and are used in the United States that focus on four topical areas: "housing and neighbourhood development; workforce and economic development; health, wellness, and family services; and household financial stability" (Divingri & St. Clair, 2017: 1). This survey is used to explain and account for barriers to upward social mobility and access to employment (Divingri & St. Clair, 2017). In the QASES community outlook sub-scale similar variables are observed. However, the way of executing QASES differs significantly to the standardised measures of surveys or questionnaires, which requires individual or household participation.

QASES also incorporates a subjective SES approach with the types of qualitative research activities employed. 'Subjective' measures of SES refer to engaging with the research participants for them to provide their own perspectives and perceptions of their communities' SES (Howe et al., 2011). The tools in QASES for gaining subjective approaches are gathered using observations, interviews and group discussions. This information from the participants help to guide the researchers when ascribing SES using QASES. Subjective approaches from the community members and participants is a way to direct the associations of the researchers' subjective decision-making when ascribing SES using the QASES scale. Moreover, QASES uses the subjective ascription of a small number of researchers to decide on SES, and is not determined directly from the community residents themselves. This is in contrast to 'objective' measures where the individual or household's SES are determined by the researcher based on the data (Howe et al., 2011).

4.3. The efficiency of the QASES methodology used to evaluate SES versus a traditional, individual-level composite-measure survey

With the efficiency analysis, I aimed to evaluate the cost-effectiveness, in terms of labour requirements and time-effectiveness, of the QASES method compared to a standard survey method. I specifically evaluated the data collection processes in order to determine the types of research activities and the labour force requirements. By looking at the labour requirements and research activities, I could determine estimates of how much it would cost to employ workers and how long it would take to complete the activities. I applied both methods to a small study sample comprising of 1 community with 8 neighbourhoods and a large study sample of a city comprising of 100 study communities with 18 neighbourhoods each. The scales of the study samples allowed me to compare the resource requirements in costs and time when applying both methods.

In both of these samples I have found that applying QASES is more cost- and time-effective than the standard SES survey approach (see Table 6 and Table 8). I want to further address this finding from what is known from the literature on the challenges of standard surveys. According to Krosnick, Presser, Fealing and Ruggles (2015: 18) there continues to be an increase in the need to obtain information on the "public's behaviours, experiences, needs and preferences". However, to conduct such high-quality surveys are becoming ever more challenging. Firstly, the cost of collecting survey data is increasing (Couper, 2017). One of the reasons are that non-response and non-coverage biases are contributing to the increased costs of individual-level methods (Couper, 2017). In the case of collecting data in LMICs, data collectors must attempt repeated visits to households not initially reached. Due to the increased efforts to reach all randomly selected households to increase response rates, grant funding for large-scale surveys has been in decline (Couper, 2017). Ultimately, response rates are regarded as proxies for data collection quality. These factors increase the costs associated with large-scale survey data collection while consequently reducing the value of surveys (Couper, 2017). On the other hand, QASES data collection does not include the challenge of reaching optimal response rates as it is a once-off data collection approach, making QASES more time- and cost-efficient.

Secondly, standard SES indices or measures are often critiqued as limited to only a few variables that are dependent on the specified research question (Yakubu & Schutte, 2019). The issue is that there is not only one way of constructing a measure for SES that are used in standard surveys. This creates partial constructs of a multi-dimensional phenomena being studied, be it SES, poverty or inequality (Kabudula *et al.*, 2017). The literature on SES constructs tends to focus on a sub-set of categories (e.g., traditional indicators of SES or wealth index) that are implemented differently for each study (Duncan, Daly, McDonough & Williams, 2002; Howe *et al.*, 2011). In contrast, the hypothetical experiments in the efficiency analysis allowed me to illustrate how QASES can be adapted to incorporate a wider spectrum of SES focus areas for a research question. That is, QASES can yield information that cannot typically be obtained from quantitative surveys through the use of qualitative tools (e.g., observations, participatory discussions and interviews).

4.4. The accuracy of the QASES method relative to a gold standard measure (PC0)

Even though QASES correlate significantly well with PC0 SES, some aspects of the QASES measure impair the strength of associations. Firstly, when observing the strength of associations between the two measures, I have found that the QASES sub-scale of assets had the least significant correlation to PC0 SES. This is observed in all the correlations between QASES and PC0 SES – assets as an individual sub-scale, and assets in combination with the other sub-scales (see Table 10, Table 12, and Table 14). Even when correlations were run between QASES sub-scales, assets delivered the weakest strength of associations (see Table 6 and Table 8). The weak association between QASES assets and PC0 SES may be attributed to the overestimation of the QASES tool when it is used in the field. The QASES tool gives a general estimate of the amount of assets per household in a zone, judging by the types of housing and community outlook that was observed during BBS data collection. The judgement among researchers when determining a score for assets was based on the common housing structures that are generally found in a zone. To assume that a well-built house has a significant amount of assets can be misleading in a country like South Africa. The delivery of basic housing to the poor is a priority to the government and has gained more traction than employment creation (The National Department of Housing, 2007). Housing recipients struggle

to maintain their houses that consequently leads to not having or owning many assets (Fieuw & Mitlin, 2018). On the contrary, in the PC0 questionnaire, the exact amount of assets is noted per household giving an accurate representation of assets distribution per community zone. Another explanation is that the assets selected for both scales are not identical (See Appendix A and Appendix B), and therefore, the construction and use of different asset indices are not directly comparable (Brodish, 2015).

Secondly, the QASES sub-scale of community outlook yielded surprisingly significant correlations with PC0 SES (r = .624). The reason why this is unanticipated, was that upon initial investigation of the scale variables for both measures, I could not directly link any attributes of the community outlook sub-scale to PC0 SES variables. As seen in Appendix A of the QASES scale, community outlook does not look at observable features. Rather it involves scenarios where a sense of a community is described in terms of social standing, social mobility, opportunities, aesthetic circumstances and safety. In the PC0 questionnaire exact measurable attributes are noted, i.e., how much dependency on government grants, type of employment, food shortages, etc. (See Appendix B). However, these factors in PC0 encompass the standard of living in a household and this might be where the two measures connect. Measures of income and expenditure influence the general well-being of an individual or household, which equally impacts how people would rate the community that they live in based on economic and social prosperity or lack thereof (Marmot, 2002; Mayosi & Benatar, 2014).

As anticipated, the housing sub-scale of QASES delivered the most significant association with PC0 SES (r = .768). This is because type of housing structures is easy to observe and classified looking at building materials, structure size, maintenance and security. Both measures had very similar approaches to classifying housing and this comes from the standardised approach of the wealth index that are commonly used in LMICs (Howe *et al.*, 2012; Kabudula *et al.*, 2017). Also, in the case of South Africa, housing policy has been a main priority for the government by means of meeting the basic needs of the population (The National Department of Housing, 2007). A great deal of resources has gone into providing subsidised housing to the most poor and vulnerable (for instance,

housing policies like the Reconstruction and Development Programme and Breaking New Ground) (Lall, Van den Brink, Dagupta & Leresche, 2012). These housing projects are often located nearby historical townships, on the edge of urban areas far from urban centres (Department of Human Settlements, 2019). This contributes to a dislocation of housing from economic opportunities that reinforce geographical and racial distortions enforced by apartheid (Fieuw & Mitlin, 2018).

The layout of housing in an area or community can be divided according to housing types – RDP housing, informal dwellings, subsidised housing apartments, privately-owned well-maintained or deteriorated houses. All these different housing can be found simultaneously within a community, usually divided along their specific types, or a mixture of different types (e.g., free-standing RDP houses with backyard dwellings). This can be considered as intra-class divisions, meaning that different levels of SES persist in a community that are conceived as low-income (Pieterse, 2009). Many have received government housing over the past few decades, despite a growing backlog (Abimbola, 2013). However, stark economic decline has led to people being unable to maintain their houses adequately (Abimbola, 2013; Fieuw & Mitlin, 2018). Therefore, housing can be an ambiguous variable to measure in South Africa where a built structure does not necessarily reflect better SES than an informal dwelling (Western Cape Governement, 2015). It becomes even more complex when determining durable household assets as an indicator of wealth when looking at housing structures.

4.5. Alternative explanations of findings

4.5.1. Different cross-sectional data

The data for both measures were collected around the same time period (2013-2015) in the exact same study communities, therefore making my study cross-sectional. However, QASES are considered as a pre-trial method, meaning that data are collected to inform a clinical trial (in this case, HPTN 071 (PopART)). When QASES data was collected in 2013, SES might have changed when PC0 commenced in 2014. The differences in cross-sectional data may have had an influence on the strength of associations between the two measures. This is in terms of housing upgrades or deteriorations as well as additional informal settlements/dwellings in a community zone. Because the

data for both QASES and PC0 were not collected at the exact same time, this might have had an influence on how well the two measures compare and correlate. This can be considered a limitation of cross-sectional data (Rafferty, Walthery & King-Hele, 2015).

4.5.2. The occurrence and meaning of outliers

Figure 12 in the findings section graphically illustrates the scatter plot data of PC0 SES QASES total. The significant finding from this graph is that there were 3 outliers present when the two datasets were correlated. I explained in sections 2.11.1.1 and 2.11.1.3 that these outliers may be as a consequence of two datasets not overlapping correctly after alterations were made. According to Danganan and Sison (2018) the overlapping of zones can create noise and inconsistent patterns, otherwise known as outliers.

In Figure 11, the bar graph of the PC0 scores also indicate few scores on the left that represents the extreme SES scores. These three scores are zones that have the lowest SES levels out of 145 zones and are considered outliers in the PC0 data. It might be a possibility that too few household questionnaires were completed for those specific zones, which understate the overall SES score of a zone. This is typically a problem when conducting studies using household surveys, where either the respondent provides inaccurate information, or the requested information is not provided at all (Wolff, 2001). When households were sampled for the PC, there were failed attempts at certain households influencing a lack of data due to nonresponse rate. The PC0 was dependent on a high response rate from the participants. If more households were ideally recruited, the results could have potentially caused less bias. From the PC data collection, not all households could be recruited due to loss to follow up of participants, ineligible selected participants, and non-response rate when visits were unsuccessfully attempted.

Furthermore, the participants that consented for PC had to fall within a specific age limit of 18-45 and were mostly women (Hayes *et al.*, 2019). A possible reason why men are underrepresented in the PC are due to men being away from home or at work during the day while women perform household or informal labour (Miraftab, 2010). This may have led to inaccurate or skewed information from the women who were selected to participate. Men are in many instances still the

main income bearers in the families by either earning the highest income or having formal employment (Chant, 2013; Statistics South Africa, 2013a). The fact that more women were appointed as field staff for the PC also explains why most participants were women in the study. This is critical when looking at this through the gender lens of health workers being predominantly women and how this impact success rates of recruiting men in health studies (Morgan, Ayiasi, Barman, Buzuzi, Ssemugabo, Ezumah, George, Hawkins, Hao, King, Liu, Molyneux, Muraya, Musoke, Nyamhanga, Ros, Tani, Theobald, Vong & Waldman, 2018). These mentioned factors limit the generalisation of the findings that were yielded from PC0.

4.5.3. Challenges with the PC0 questionnaire

Upon investigation of the correlations between QASES and PC0, I also explored the possibility that the gold standard SES measurement (PC0 SES) may have inherent flaws. I investigated the PC0 questionnaire in order to compare the types of variables that are used to determine SES. This speaks to what is highlighted in research: some variables need to be redefined to indicate SES more accurately (Milenkovic *et al.*, 2014). I attempted to divide the PC0 questionnaire up into different sections that are encapsulated as variables. These are housing characteristics (floor, roof and structure material) and amenities (type of toilet and electricity), income/expenditure (employment type, dependency on grants, food security and spending on necessities), and assets (TV, refrigerator, stove, radio, etc.). Together, these sections or variables in the questionnaire make up of what is termed a wealth index that was originally constructed from existing data on household assets, amenities, and services in order to structure indicators such as health and population according to economic status (Rutstein, 2004).

4.5.4. Discussion of PC variables

The Economic and Food security section in the PC0 questionnaire (Appendix B) is too broad in its questions, meaning that the type of criteria does not fit the South African context well enough. It focuses on a broader African context. For instance, the Western Cape, specifically in the Cape metropolitan and Cape Winelands areas, are common for having Wendy (wooden structure) houses or

multiple backyard dwellings on one plot (Govender, 2011; Western Cape Governement, 2015: 8–9). In the questionnaire there is no option for the participant to be asked to identify their housing structure. In the Western Cape, the delivery of subsidised housing is unable to keep up with a growing demand for housing that are fuelled by rapid urbanisation and migration in Cape Town (Department of Human Settlements, 2019). Therefore, many people continue to settle in overcrowded houses, informal houses and backyard dwellings (Department of Human Settlements, 2019). Also, the circumstances may be observationally/physically better for South Africans who mostly have access to electricity and flush toilets, but the assumption cannot be made that these are strong determinants of better SES. The Human Settlement Demand Profile (Western Cape Governement, 2015) indicate that there are higher-income households (>R30 000) that live in informal dwellings, whilst households with brick or concrete structured houses contained the majority of the poor (159,019 out of 344,551) living in the lowest monthly income bracket (R0-R800). Providing large-scale housing projects (RDP and BNG) in the form of free-standing housing units, is one of the most prominent governmental approaches to reparation after apartheid was demolished. Therefore, it is problematic to assume that basic housing, amenities and services in the PC0 questionnaire equates to better SES when the Western Cape economy were expected to grow by only 1% in 2018 (Department of Human Settlements, 2019). Ultimately, the use of standardised wealth indices can have many limitations for specific contextual factors that are unique to a country or even a community/neighbourhood.

The same can be said for the income/expenditure variable. For example, the types of employment are not very representative to the context. In the case of farming, not many people rely on subsistence and smallholder farming as a means of income in the city/metropolitan areas of Cape Town (Statistics South Africa, 2013b). Out of all the provinces, the Western Cape has the highest proportion of agricultural commercial farming intended for the market (Statistics South Africa, 2013b). Within a 75km radius of Cape Town, most of the commercial agricultural output is in "intensive poultry, pork, vegetable, and milk production" (Vink & Tregurtha, 2005: 3). Therefore, people would more commonly be employed as farm workers working on commercial agricultural farms in the surrounds of cities and towns, than depend on small-scale farming. This view is

supported by the fact that 232 000 workers were employed in agriculture in the Western Cape, which is the second largest employment sector after manufacturing (Ungerer, Bowmaker-Falconer, Oosthuizen, Phehane & Strever, 2018). Fishing, as a large agricultural sector in the greater Cape Town region (Vink & Tregurtha, 2005), is also not very applicable to the specific study communities of HPTN 071 (PopART), as most study areas are located far from fishing locations except for two study communities in Khayelitsha.

As already highlighted, in LMICs an asset index, which is constructed from a list of household asset items, is one of the most widely used measures of SES (Ataguba, Akazili & McIntyre, 2011; Kabudula *et al.*, 2017). The preference for the use of an asset index are due to time and cost-effectiveness (Kabudula *et al.*, 2017). The increasing use of an asset index can also be attributed to unavailability of expenditure data and difficulty of collecting such data (Filmer & Scott, 2012). However, these benefits also come with limitations. According to Ewerling and Barros (2017), the asset index does not account for rapid income and technology improvements that popularise expensive assets in a short amount of time. It is therefore recommended that asset indicators require periodic updates to avoid losing strength of an indicator (Ewerling & Barros, 2017). For instance, certain assets must be avoided, such as radios and DVDs, as fewer people happen to own such equipment. Additionally, other equipment is considered as long-lasting indicators, such as TVs, computers/laptops or internet. Number of TVs or computers per household may correlate high with wealth, therefore increasing the ability to differentiate between higher and lower income households (Ewerling & Barros, 2017). In the PC0 questionnaire, some asset items can be regarded as outdated, e.g., a stereo/cassette player and radio. Most people would not own a stereo/cassette player or radio anymore due to it being outmoded. This might affect the number of ticks obtained from the scale seeing that the more assets you have the better off you are, which also varies in terms of the value of each asset. Additionally, the list of assets is very limited as it was the goal not to lengthen the questionnaire any further. The basic asset items listed, might not be sufficiently relevant to current technological developments.

4.6. Strengths of my study

The main strength of this study is that I was equipped with a large sample size that enabled me to do a detailed exploration of correlations across the data. Through these explorations, I could determine how a simplified SES measurement, without the burden of costs, time and long questionnaires, can yield valid results of a community's SES. Additionally, I could do a thorough investigation of the conceptualisation of SES, specifically the contextual analysis of SES. I explored indicators/variables of SES and how they are applied in health-related research. I have gathered knowledge on the implications of using narrow descriptions in standardised SES scales to promote cross-country comparisons.

The secondary data that I used for my analyses are of good quality, with regards to data collection and expertise involved. A great amount of people from various backgrounds and knowledge were involved in both the BBS and PC studies. As thoroughly explained in Chapter 2: Method, both studies followed clear cut and pre-specified sets of methods that supported and steered the research outputs of HPTN 071 (PopART). I can confidently state that the generated data yielded quality results from my secondary analyses, which is a strength of this study.

From the results, the 99% confidence level of the correlations implies that the findings can be generalised to other contexts. This is true irrespective of the different data types (ordinal, interval and cross-sectional) and non-parametric statistical tests applied. The large number of study communities and their locations in two different districts ensured heterogeneity. Random sampling was applied to the study populations and communities, therefore increasing the likelihood of the probability that similar outcomes can be observed elsewhere (Polit & Beck, 2010). Specifically, related results might be found in communities that are like the study communities used in this study. I would advise further exploration in other settings to refine generalisability.

4.7. Limitations to my findings

I was not involved in data collection for BBS or PC, and therefore I do not have the advantage as a researcher to know what took place in reality during fieldwork regarding specific problems that were identified, or how QASES was ascribed. However, I had access to many people working on the

studies as well as various papers and reports produced on the methodologies applied (for instance Hayes, Bond, Hargreaves etc.).

The QASES sub-scale of assets produced the least significant correlations to the PC0 data. It is evident that it is difficult to judge the use of the assets sub-scale by only observing the types of houses and approaching people on the street to have short conversations about their durable assets. The QASES process does not adequately represent the complex nature of household assets by only doing observational activities or a few interviews and group discussions. If assets should be measured in QASES going forward, other methods for doing so need to be constructed to improve its accuracy.

The data that were compared and correlated are inherently different. That is, QASES data were captured as ordinal values, while PC0 data were captured as interval values. To apply correlation analysis became tricky when dealing with different data types. This means that to analyse the association of the data using Spearman's correlation coefficient, are less sensitive to strong outliers found in the tails of both data samples (at the extreme ends). This limits the detection of bias in the data. Additionally, Spearman's correlation indicates covariance but does not imply that a causal relationship exists (Allen, 2017).

The use of different cross-sectional data also entails certain weaknesses. For instance, each dataset was collected separately at approximately the same timeframe (BBS in 2013 and PC0 in 2014). These 'snapshots' do not provide an indication that differed results would have been found if a different time point had been chosen. Therefore, cross-sectional studies do not provide precise information about causal relationships (Allen, 2017). Because the data of QASES were collected at a specific point compared to data collected for PC at another time, bias in the correlations might be present.

4.8. Pragmatic recommendations about the use of the QASES method for policy and practice

4.8.1. The use of QASES as a neighbourhood/community SES measure

According to Fotso and Kuate-Defo (2005), growing evidence point towards the importance of context on health inequalities and, therefore, it is necessary to focus on the characteristics of the

community or neighbourhood where individuals or households reside. With this fact in mind, QASES is an observational tool that measures SES on a neighbourhood level, whereby interactions with community members are valuable to the process. Instead of measuring SES on the individual level with the use of a questionnaire or survey (PC0), QASES are done rapidly by observing the contextual factors of a community. This method adds to the existence of socio-economic diversity within a community, which exemplifies heterogeneity not only across communities, but also within. This is an important contribution to policy that are oftentimes too devoted to large-scale survey evidence and neglecting contextual variables and differences (Hannscott, 2015; Hantrais, Lenihan & MacGregor, 2015). Improvement of the practice should be directed towards developing SES measurements that are context specific.

4.8.2. Improvement of QASES measurement for policy implementation

In order to apply the QASES method elsewhere, the following recommendations can be considered to enable better practice and increase validity of data:

The sub-scale of housing yielded the most significant associations overall. Alternatively, the sub-scale of assets yielded the least significant associations, even though still producing moderately significant results. I would suggest that mathematical adjustments can be made after the scores are ascribed according to each sub-scale. From these results, weighting down assets and weighting up housing can account for an improved SES estimation since housing are easier to observe and measure. Additionally, I would also recommend that the QASES method should be applied and explored in other contexts in South Africa and elsewhere to develop the method's broader applicability. The method's approach can be adapted to suit the resource availability when applied elsewhere as illustrated in the efficiency analysis (section 3.2.).

4.8.3. Policy recommendations for standardised SES measurement

The case for using an asset index should be developed properly based on specific contextual details. From the research findings and indicated literature, a universal asset index was used for PC0 that was intended to be applicable across LMICs, or more specifically the African context. However,

there exist so much variation in housing structures and assets (combined as the wealth index) across countries, that the model cannot be universally applied. I suggest that a carefully considered asset index are constructed when selecting the list of household assets for each study context when measuring SES.

The same principle applies for the housing variable. In the discussion of findings, the issue was highlighted that there is no mention of 'backyard dwelling' as housing category. A category for cluster/multi-units are included in the questionnaire, which refers to multiple structures on one plot (e.g., semi-detached brick/wood/corrugated iron sheet structures). This can be used as a category for backyard dwellings, but it does not encapsulate the precise circumstances of housing dynamics in South Africa. I suggest that 'backyard dwellings' be included as a housing category in standardised questionnaires.

With random selection of households to complete questionnaires, it should be clearly stated how many households in a zone did participate to determine the significance of an aggregate result. As in the case with the PC0 data, there were accounts of zones where only a few households participated, which may explain the outliers in the data that represents a biased/inaccurate score. For the implementation of standardised SES measurement where households or individuals are randomly selected, more detail should be provided to explain how many households participated and how this influences the study results, especially in an instance of a small sample in a zone.

4.9. Future research opportunities

For standard SES measurement, research opportunities are towards updating and developing SES indicators to fit the South African context. The focus should be to construct a contemporary asset index that excludes non-applicable technologies (e.g., cassette player or radio) and includes up-to-date technologies like the internet. The same principle is applied to updating housing categories (e.g., including backyard dwellings as a category). More research should also focus on determining the major employment sectors that are context specific so that questionnaire options can be updated for

employment categories. These research opportunities are raised so that inclusivity is promoted when conducting research in previously disadvantaged communities.

Instead of conducting correlation analysis to improve the QASES method, it is also recommended to perform statistical analysis that compare the two measurement techniques (e.g., Bland-Altman test) (Ranganathan, Pramesh & Aggarwal, 2017). The Bland-Altman test can analyse the level of agreement between the QASES and PC0 SES methods with the use of plots (Hanneman, 2008). This type of analysis can determine whether QASES can be used as a substitute for the standard PC0 SES scale. Both data should be organised on interval scales to perform the Bland-Altman test (Ranganathan *et al.*, 2017). Statistical tests of agreement are also available to assess inter-rater variability between measurements of categorical variables (e.g., Weighted kappa) (Ranganathan *et al.*, 2017).

I would also recommend that further studies are conducted where QASES are applied to different study samples in conjunction with a standard SES measure. The two methods should be performed simultaneously in order to compare or correlate the same cross-sectional data. It would be particularly interesting to apply the QASES method in high-income communities so that cross-comparison between communities of different SES can occur. This is a way to increase the probability of generalisability.

4.10. Conclusion

In this research I explored the notion of using a qualitative tool (QASES) to measure SES. This is different from conventional SES measurement – which is done typically using a questionnaire applied at a large scale (community/neighbourhood, province or country). The significance and contribution of SES to health has to do with providing evidence that individuals with the lowest levels of SES have the highest disease and death rates. Therefore, the priority persists to develop SES in order to inform health policy where the most need for resource implementation exist.

In the literature there is a lot of variation in how SES is measured either as an objective measure or subjective measure applied at different levels (individual, household or neighbourhood) and different approaches (e.g., compositional, area/contextual, composite or multilevel). The effort has always been towards creating SES indicators that are measurable across different contexts – be it within countries or across countries. This has to do with standardising SES to create opportunities for cross-comparisons. However, there has been many inadequacies and limitations with this approach, especially with how SES is measured in HICs compared to LMICs. It is evident that the way SES is measured in high-income countries is standardised – collecting data at a large-scale using questionnaires where traditional indicators of SES are used (income, education and occupation). This approach has also been standardised in LMICs, while different indicators are incorporated into a scale or measure such as consumption/expenditure, education, occupation and a combination of indicators that form part of a wealth index – assets, housing characteristics. From these different approaches to measuring SES, cross-comparison between countries become complex as the same indicators are not applied universally.

I explored the efficiency and accuracy of QASES by comparing data collected using the QASES measure to data collected using the standard SES measure in the PC0. The process of executing QASES is much more efficient in terms of time and financial costs. Compared to the standard way of SES measurement where a large labour force and substantial time are required. The difference is that surveys have added costs to reach response rates by repeated household visit attempts, while QASES encompass once-off data collection activities. Furthermore, I have found that

the data between the two measures correlate at significant levels (e.g., correlation between QASES total and PC0 at r=.753). I have also determined that the QASES measure scores relatively similar values compared to the PC0 SES measure. Therefore, I conclude that QASES is a quicker, cost-effective and operationally useful measure to implement for health intervention studies and policy planning.

The relevance of my study was to demonstrate that the implementation of SES in research can be done in a quick and operationally-friendly manner. I have described that measuring SES with a rapid and qualitative approach can yield significant results that get close enough to results of SES measured in the standard way. What can also be gained from the results and literature study is that too much emphasis is placed on adopting strategies for measuring SES that are standardised to allow for cross-comparisons between countries or provinces. The question going forward is to determine how relevant this notion is when there is a prominent shift towards the regard for contextual differences. Instead, it is essential to improve the practice of measuring SES based on the differences in contexts and to develop indicators and scales as such.

References

Abimbola, O.W. 2013. Value-based perspectives of stakeholders' building requirements in low cost and government subsidised housing projects in South Africa. *Construction Innovation*. 13(4):424–444.

Adler, N.E., Epel, E., Castellazzo, G. & Ickovics, J. 2000. Relationship of subjective and objective social status with psychological and physical health: preliminary data in healthy white women. *Health Psychology*. 19:586–592.

Aggarwal, O.P., Bhasin, S.K., Sharma, a K., Chhabra, P., Aggarwal, K. & Rajoura, O.P. 2005. A New Instrument (Scale) for Measuring the Socioeconomic Status of a Family : Preliminary Study. *Indian Journal of Community Medicine*. 30(4):3–6. [Online], Available: http://books.google.com/books?id=vvkceVbJlSkC&pgis=1.

Allen, M. 2017. *The SAGE Encyclopedia of Communication Research Methods*. Thousand Oaks, California.

Allen, I.E. & Seaman, C.A. 2007. Likert scales and data analyses. *Quality Progress*.

Ataguba, J.E., Akazili, J. & McIntyre, D. 2011. Socioeconomic-related health inequality in South Africa: Evidence from General Household Surveys. *International Journal for Equity in Health*. 10(1):48.

Bärnighausen, T., Hosegood, V., Timaeus, I.M. & Newell, M.-L. 2007. The socioeconomic determinants of HIV incidence: evidence from a longitudinal, population-based study in rural South Africa. *AIDS*. 21(Suppl 7):S29–S38.

Beegle, K., De Weerdt, J., Friedman, J. & Gibson, J. 2012. Methods of household consumption measurement through surveys: Experimental results from Tanzania. *Journal of Development Economics*. 98(1):3–18.

Benatar, S. 2013. The challenges of health disparities in South Africa. *S Afr Med J*. 103(3):154–155.

Berger, N. & Archer, J. 2018. Qualitative insights into the relationship between socioeconomic status

and students' academic achievement goals. *Social Psychology of Education*. 21(4):787–803.

Bernheim, S.M., Ross, J.S., Krumholz, H.M. & Bradley, E.H. 2008. Influence of patients' socioeconomic status on clinical management decisions: A qualitative study. *Annals of Family Medicine*. 6(1):53–59.

Berzofsky, M., Smiley, H. & Krebs, C. 2014. *Measuring Socioeconomic Status (SES) in the NCVS: Background , Options , and Recommendations*.

Bond, V., Hoddinott, G., Musheke, M., Viljoen, L., Abrahams, K., Chiti, B., Mantantana, J., Ndubani, R., et al. 2013. *Broad brush surveys of HIV prevention, treatment and care in 21 Zambian and South African communites to prepare for HPTN071 (PopART).*

Bond, V., Chiti, B., Hoddinott, G., Reynolds, L., Schaap, A., Simuyaba, M., Ndubani, R., Viljoen, L., et al. 2016. "The difference that makes a difference''': highlighting the role of variable contexts within an HIV Prevention Community Randomised Trial (HPTN 071/PopART) in 21 study communities in Zambia and South Africa". *Aids Care-Psychological and Socio-Medical Aspects of Aids/Hiv*. 28(3, SI):99–107.

Bond, V., Ngwenya, F., Murray, E., Ngwenya, N., Viljoen, L., Gumede, D., Bwalya, C., Mantantana, J., et al. 2018. Value and Limitations of Broad Brush Surveys Used in Community-Randomized Trials in Southern Africa. *Qualitative Health Research*. 104973231880994.

Braveman, P.A., Cubbin, C., Egerter, S., Chideya, S., Marchi, K.S., Metzler, M. & Posner, S. 2005. Socioeconomic Status in Health Research. *JAMA*. 294(22):2879.

Brodish, P.H. 2015. An association between neighbourhood wealth inequality and HIV prevalence in sub-Saharan Africa. *Journal of Biosocial Science*. 47(3):311–328.

Bunyasi, E.W. & Coetzee, D.J. 2017. Relationship between socioeconomic status and HIV infection: findings from a survey in the Free State and Western Cape Provinces of South Africa. *BMJ open*. 7(11):e016232.

Chant, S. 2013. Cities through a "gender lens": A golden "urban age" for women in the global South?

Environment and Urbanization. 25(1):9–29.

Coovadia, H., Jewkes, R., Barron, P., Sanders, D. & McIntyre, D. 2009. The health and health system of South Africa: historical roots of current public health challenges. *The Lancet*. 374(9692):817–834.

Couper, M.P. 2017. New Developments in Survey Data Collection. *Annual Review of Sociology*. 43(1):121–145.

Cowan, C.D., Hauser, R.M., Kominski, R. a., Levin, H.M., Lucas, S.R., Morgan, S.L., Spencer, M.B. & Chapman, C. 2013. Improving the Measurement of Socioeconomic Status for the National Assessment of Educational Progress: A Theoretical Foundation. *Recommendations to the National Center for Education Statistics*. (November):1–37. [Online], Available: http://nces.ed.gov/nationsreportcard/pdf/researchcenter/Socioeconomic_Factors.pdf.

Cutler, D.M., Lleras-Muney, A. & Vogl, T. 2008. *Socioeconomic Status and Health: Dimensions and Mechanisms*. Cambridge.

Danganan, A.E., Sison, A.M. & Medina, R.P. 2018. An improved overlapping clustering algorithm to detect outlier. *Indonesian Journal of Electrical Engineering and Informatics*. 6(4):401–409.

Department of Human Settlements. 2019. *Living Cape: A Human Settlements Framework.*

Diniz, E., Castro, P., Bousfield, A. & Figueira Bernardes, S. 2020. Classism and dehumanization in chronic pain: A qualitative study of nurses' inferences about women of different socio-economic status. *British Journal of Health Psychology*. 25(1):152–170.

Divingri, E. & St. Clair, N. 2017. *Community Outlook Survey: Barriers to Employment Access Remain Key Challenge in Workforce and Economic Development.* [Online], Available: https://www.philadelphiafed.org/community-development/community-outlook-survey.

Duncan, G.J., Daly, M.C., McDonough, P. & Williams, D.R. 2002. Optimal indicators of socioeconomic status for health research. *American Journal of Public Health*. 92(7):1151–1157.

Ewerling, F. & Barros, A.J.D. 2017. After 10 years, how do changes in asset ownership affect the Indicador Econômico Nacional? *Revista de saude publica*. 51:10.

Eyre, E.L.J., Duncan, M.J., Birch, S.L. & Cox, V.M. 2014. Low socio-economic environmental determinants of children's physical activity in Coventry, UK: A Qualitative study in parents. *Preventive Medicine Reports*. 1:32–42.

Field, A. 2009. *Discovering Statistics using SPSS*. Third edit ed. London: SAGE Publications Ltd.

Fieuw, W. & Mitlin, D. 2018. What the experiences of South Africa's mass housing programme teach us about the contribution of civil society to policy and programme reform. *Environment and Urbanization*. 30(1):215–232.

Filmer, D. & Scott, K. 2012. Assessing Asset Indices. *Demography*. 49(1):359–392.

Finchilescu, G. & Tredoux, C. 2010. The Changing Landscape of Intergroup Relations in South Africa. *Journal of Social Issues*. 66(2):223–236.

Fliesser, M., De Witt Huberts, J. & Wippert, P.M. 2018. Education, job position, income or multidimensional indices? Associations between different socioeconomic status indicators and chronic low back pain in a German sample: A longitudinal field study. *BMJ Open*. 8(4):1–7.

Fotso, J.-C. & Kuate-Defo, B. 2005. Measuring socioeconomic status in health research in developing countries: Should we be focusing on households, communities or both? *Social Indicators Research*. 72(2):189–237.

Galobardes, B., Shaw, M., Lawlor, D.A., Lynch, J.W. & Smith, G.D. 2006a. Indicators of socioeconomic position (part 2). *Journal of Epidemiology and Community Health*. 60(2):95–101.

Galobardes, B., Shaw, M., Lawlor, D.A., Lynch, J.W. & Smith, G.D. 2006b. Indicators of socioeconomic position (part 1). *Journal of Epidemiology and Community Health*. 60(1):7–12.

Galobardes, B., Lynch, J. & Smith, G.D. 2007. Measuring socioeconomic position in health research. *British Medical Bulletin*. 81–82(1):21–37.

Garth, A. 2008. *Analysing data using SPSS.* [Online], Available: https://students.shu.ac.uk/lits/it/documents/pdf/analysing_data_using_spss.pdf.

Gaur, K.L. 2013. Socio-Economic Status Measurement Scale: Thirst Area With Changing Concept For Socio-Economic Status. *International Journal of Innovative Research and ….* 2(9):139–145. [Online], Available: http://www.ijird.com/index.php/ijird/article/view/37303.

Govender, T. 2011. The health and sanitation status of specific low-cost housing communities as contrasted with those occupying backyard dwellings in the City of Cape Town, South Africa. Stellenbosch University. [Online], Available: http://hdl.handle.net/10019.1/17992.

Grant, T.L., Edwards, N., Sveistrup, H., Andrew, C. & Egan, M. 2010. Inequitable walking conditions among older people: Examining the interrelationship of neighbourhood socio-economic status and urban form using a comparative case study. *BMC Public Health*. 10.

Gutreuter, S., Igumbor, E., Wabiri, N., Desai, M. & Durand, L. 2019. Improving estimates of district HIV prevalence and burden in South Africa using small area estimation techniques. *PLoS ONE*. 14(2):1–14.

Hanneman, S.K. 2008. Design, analysis, and interpretation of method-comparison studies. *AACN Advanced Critical Care*. 19(2):223–234.

Hannscott, L. 2015. Individual and contextual socioeconomic status and community satisfaction. *Urban Studies*. 53(8):1727–1744.

Hantrais, L., Lenihan, A.T. & MacGregor, S. 2015. Evidence-based policy: exploring international and interdisciplinary insights. *Contemporary Social Science*. 10(2):101–113.

Hayes, R. & Bennett, S. 1999. Such As School Climate, Cultural Attitudes and Beliefs, Poverty Status and Community Environment. *International Journal of Epidemiology*. 28:319–326.

Hayes, R. & Fidler, S. 2015. *HPTN 071 Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa.*

Hayes, R., Ayles, H., Beyers, N., Sabapathy, K., Floyd, S., Shanaube, K., Bock, P., Griffith, S., et al. 2014. HPTN 071 (PopART): Rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including universal testing and treatment - a study protocol for a cluster randomised trial. *Trials*. 15(1):1–17.

Hayes, R., Ayles, H., Beyers, N. & Sabapathy, K. 2014. HPTN 071 (PopART): rationale and design of a cluster-randomised trial of the population impact of an HIV combination prevention intervention including. *Trials*. 1–17. [Online], Available: http://www.biomedcentral.com/content/pdf/1745-6215-15-57.pdf.

Hayes, R.J., Donnell, D., Floyd, S., Mandla, N., Bwalya, J., Sabapathy, K., Yang, B., Phiri, M., et al. 2019. Effect of universal testing and treatment on HIV incidence - HPTN 071 (PopART). *New England Journal of Medicine*. 381(3):207–218.

Howe, L.D., Hargreaves, J.R. & Huttly, S.R.A. 2008. Issues in the construction of wealth indices for the measurement of socio-economic position in low-income countries. *Emerging Themes in Epidemiology*. 5:1–14.

Howe, L.D., Hargreaves, J.R., Ploubidis, G.B., De Stavola, B.L. & Huttly, S.R.A. 2011. Subjective measures of socio-economic position and the wealth index: A comparative analysis. *Health Policy and Planning*. 26(3):223–232.

Howe, L.D., Galobardes, B., Matijasevich, A., Gordon, D., Johnston, D., Onwujekwe, O., Patel, R., Webb, E.A., et al. 2012. Measuring socio-economic position for epidemiological studies in low- and middle-income countries: A methods of measurement in epidemiology paper. *International Journal of Epidemiology*. 41(3):871–886.

Ichoku, H. 2011. On the Use of Socioeconomic Status Indicators in the Analysis of Health Inequalities and Poverty in Africa. *Journal of International Development*. 23:1004–1012.

Jabar, A. & Matzopoulos, R. 2017. Violence and injury observatories: Reducing the burden of injury in high-risk communities. *South African Crime Quarterly*. (59):47–57.

Kabudula, C.W., Houle, B., Collinson, M.A., Kahn, K., Tollman, S. & Clark, S. 2017. Assessing Changes in Household Socioeconomic Status in Rural South Africa, 2001–2013: A Distributional Analysis Using Household Asset Indicators. *Social Indicators Research*. 133(3):1047–1073.

Kaplan, G.A., Hann, M.N., Syme, S.L., Minkler, M. & Winkleby, M. 1987. Socioeconomic status and health. *American Journal of Preventative Medicine*. 3:125-129.

Kevany, S., Benatar, S.R. & Fleischer, T. 2013. Improving resource allocation decisions for health and HIV programmes in South Africa: Bioethical, cost-effectiveness and health diplomacy considerations. *Global Public Health*. 8(5):570–587.

Krieger, N., Williams, D.R. & Moss, N.E. 1997. Measuring Social Class in US Public Health Research: Concepts, Methodologies, and Guidelines. *Annual Review of Public Health*. 18(1):341–378.

Krosnick, J., Presser, S., Fealing, K. & Ruggles, S. 2015. *The Future of Survey Research: Challenges and Opportunities.* Arlington, VA.

Kulkarni, P., Ramesh Masthi, N. & Gangaboraiah. 2013. An exploratory study on socio economic status scales in a rural and urban setting. *Journal of Family Medicine and Primary Care*. 2(1):69.

Lall, S. V, Van den Brink, R., Dagupta, B. & Leresche, K.M. 2012. *Shelter from the Storm — but Disconnected from Jobs: Lessons from Urban South Africa on the Importance of Coordinating Housing and Transport Policies*.

LaVeist, T.A. 2005. Disentangling race and socioeconomic status: A key to understanding health inequalities. *Journal of Urban Health*. 82(SUPPL. 3).

Lewandowski, C.S.C. & Bolt, S.E. 2010. Box-and-Whisker Plot. In N. Salkind (ed.). Thousand Oaks, California: SAGE Publications, Inc. *Encyclopedia of Research Design*. [Online], Available: http://dx.doi.org/10.4135/9781412961288.n35 Print.

Loewenthal, K. & Lewis, C.A. 2001. *An Introduction to Psychological Test and Scales*.

Lombard, M. 2014. Constructing ordinary places: Place-making in urban informal settlements in

Mexico. *Progress in Planning*. 94:1–53.

Lorenz, E., Köpke, S., Pfaff, H. & Blettner, M. 2018. Cluster-Randomized Studies - part 25 of a series on evaluating scientific publications. *Dtsch Arztebl Int*. 115:163–168.

Marmot, M. 2002. The influence of income on health: Views of an epidemiologist. *Health Affairs*. 21(2):31–46.

Mayosi, B.M. & Benatar, S.R. 2014. Special report Health and Health Care in South Africa — 20 Years after Mandela. *The New England Journal of Medicine*. 371(14):1344.

Mbirimtengerenji, N.D. 2007. Is HIV/AIDS epidemic outcome of poverty in sub-Saharan Africa? *Croatian Medical Journal*. 48(5):605–617.

Milenkovic, N., Vukmirovic, J., Bulajic, M. & Radojicic, Z. 2014. A multivariate approach in measuring socio-economic development of MENA countries. *Economic Modelling*. 38:604–608.

Miller, R.J. & Yang, Y. 1997. Association Rules over Interval Data. *SIGMOD Record (ACM Special Interest Group on Management of Data)*.

Miraftab, F. 2010. Contradictions in the gender-poverty nexus: Reflections on the privatisation of social reproduction and urban informality in South African townships. *The International Handbook of Gender and Poverty: Concepts, Research, Policy*. (January 2010):644–648.

Morgan, R., Ayiasi, R.M., Barman, D., Buzuzi, S., Ssemugabo, C., Ezumah, N., George, A.S., Hawkins, K., et al. 2018. Gendered health systems: Evidence from low- and middle-income countries. *Health Research Policy and Systems*. 16(1):1–12.

Morris, R. & Carstairs, V. 1991. Which deprivation? A comparison of selected deprivation indexes. *J Public Health Med.* 13(4):318–26.

Mosavel, M., Simon, C., Van Stade, D. & Buchbinder, M. 2005. Community-based participatory research (CBPR) in South Africa: engaging multiple constituents to shape the research question. *Social science & medicine (1982)*. 61(12):2577–87.

Munyoro, I. 2018. Research Data Collection in Challenging Environments: Barriers to Studying the Performance of Zimbabwe's Parliamentary Constituency Information Centres (PCICs). *The African Journal of Information and Communication*. (21):81–95.

Nama, N. & Swartz, L. 2002. Ethical and Social Dilemmas in Community-based Controlled Trials in Situations of Poverty: A View from a South African Project. *Journal of Community and Applied Social Psychology*. 12(4):286–297.

National Department of Health. 2017. *South Africa Demographic and Health Survey 2016: Key Indicators Report*.

Oakes, J. 2008. Measuring Socioeconomic Status. *e-Source: Behavioral and Social Science Research*. 1–31.

Oakes, J.M. & Rossi, P.H. 2003. The Measurement of Socioeconomic Status in Health Research: Current Practice and steps toward a new approach. *Social Science & Medicine*. 56:769–784.

Omer, W. & Al-Hadithi, T. 2017. Developing a socioeconomic index for health research in Iraq. *Eastern Mediterranean Health Journal*. 23(10):670–677.

Pieterse, E. 2009. *Post-Apartheid Geographies in South Africa: why are urban divides so persistent*.

Poirier, M.J.P., Grépin, K.A. & Grignon, M. 2020. Approaches and Alternatives to the Wealth Index to Measure Socioeconomic Status Using Survey Data: A Critical Interpretive Synthesis. *Social Indicators Research*. 148(1).

Polit, D.F. & Beck, C.T. 2010. Generalization in quantitative and qualitative research: Myths and strategies. *International Journal of Nursing Studies*. 47(11):1451–1458.

Pollack, C.E., Chideya, S., Cubbin, C., Williams, B., Dekker, M. & Braveman, P. 2007. Should Health Studies Measure Wealth?. A Systematic Review. *American Journal of Preventive Medicine*. 33(3):250–264.

Psaki, S.R., Seidman, J.C., Miller, M., Gottlieb, M., Bhutta, Z.A., Ahmed, T., Ahmed, A.M.S., Bessong,

P., et al. 2014. Measuring socioeconomic status in multicountry studies: Results from the eight-country MAL-ED study. *Population Health Metrics*. 12(1):1–11.

Rafferty, A., Walthery, P. & King-Hele, S. 2015. *Analysing change over time: repeated cross sectional and longitudinal survey data*. [Online], Available: https://www.ukdataservice.ac.uk/media/455362/changeovertime.pdf.

Ranganathan, P., Pramesh, C.S. & Aggarwal, R. 2017. Common pitfalls in statistical analysis: Measures of agreement. *Perspectives in clinical research*. 8(4):187–191.

Roshita, A., Schubert, E. & Whittaker, M. 2012. Child-care and feeding practices of urban middle class working and non-working Indonesian mothers: A qualitative study of the socio-economic and cultural environment. *Maternal and Child Nutrition*. 8(3):299–314.

Rutstein, S.O. 2004. The DHS wealth index: Approaches for rural and urban areas. *DHS Working Papers No. 60*. (60):1–71.

Saif-Ur-Rahman, K.M., Anwar, I., Hasan, M., Hossain, S., Shafique, S., Haseen, F., Khalequzzaman, M., Rahman, A., et al. 2018. Use of indices to measure socio-economic status (SES) in South-Asian urban health studies: a scoping review. *Systematic reviews*. 7(1):196.

Salkind, C.N.J. 2010. Encyclopedia of Research Design Exploratory Factor Analysis. 459–466.

Seebregts, C.J., Zwarenstein, M., Mathews, C., Fairall, L., Flisher, A.J., Seebregts, C., Mukoma, W. & Klepp, K.I. 2009. Handheld computers for survey and trial data collection in resource-poor settings: Development and evaluation of PDACT, a Palm™ Pilot interviewing system. *International Journal of Medical Informatics*. 78(11):721–731.

Seekings, J. 2010. *Race, class and inequality in the South African City*. (283).

Shaikh, Z. & Pathak, R. 2017. Revised Kuppuswamy and B G Prasad socio-economic scales for 2016. *International Journal Of Community Medicine And Public Health*. 4(4):997.

Shavers, V.L. 2007. Measurement of socioeconomic status in health disparities research. *Journal of the*

National Medical Association. 99(9):1013–23.

Shavers, V.L. & VL, S. 2007. Measurement of socioeconomic status in health disparities research. *Journal of the National Medical Association*. 99(9):1013–23.

Shisana, O., Zungu, N. & Pezi, S. 2009. *HIV/AIDS in South Africa 25 years on: Psychosocial perspectives*. Vol. 2016. P. Rohleder, L. Swartz, S.C. Kalichman, & L.C. Simbayi (eds.).

Shisana, O., Rehle, T., Simbayi, L.C., Zuma, K., Jooste, S., Zungu, N., Labadarios, D., Onoya, D., et al. 2014. *South African National HIV Prevalence, Incidence and Behaviour Survey, 2012*.

Simpson, S.H. 2015. Creating a data analysis plan: What to consider when choosing statistics for a study. *Canadian Journal of Hospital Pharmacy*. 68(4):311–317.

Simwinga, M., Bond, V., Makola, N., Hoddinott, G., Belemu, S., White, R., Shanaube, K., Seeley, J., et al. 2016. Implementing community engagement for combination prevention: Lessons learnt from the first year of the HPTN 071 (PopART) Community-Randomized study. *Current HIV/AIDS Reports*. 13(4):194–201.

Statistics South Africa. 2013a. *Gender statistics South Africa 2011*.

Statistics South Africa. 2013b. Census 2011 Agricultural households Key Highlights Report. Report No.:3–30.

Statistics South Africa. 2015. *Living Conditions of Households in South Africa: An analysis of household expenditure and income data using the LCS 2014/2015*. [Online], Available: http://www.statssa.gov.za/publications/P0310/P03102014.pdf.

Statistics South Africa. 2016. *Community Survey 2016 in Brief*.

Sweeney, S., Vassall, A., Foster, N., Simms, V., Ilboudo, P., Kimaro, G., Mudzengi, D. & Guinness, L. 2016. Methodological Issues to Consider When Collecting Data to Estimate Poverty Impact in Economic Evaluations in Low-income and Middle-income Countries. *Health Economics (United Kingdom)*. 25:42–52.

The National Department of Housing. 2007. *Annotated Housing Bibliography 2003-2007*. Johannesburg. [Online], Available: https://www.wits.ac.za/media/wits-university/faculties-and-schools/-engineering-and-the-built-environment/research-entities/cubes/documents/Annotated Housing Bibliography 2003-2007.pdf.

Tiwari, S.C., Kumar, A. & Kumar, A. 2005. Development & standardization of a scale to measure socio-economic status in urban & rural communities in India. *Indian Journal of Medical Research*. 122(4):309–314.

UNAIDS. 2015. *UNAIDS Terminology Guidelines - 2015*. [Online], Available: https://www.unaids.org/sites/default/files/media_asset/2015_terminology_guidelines_en.pdf.

UNAIDS. 2018. *HIV Prevalence*. [Online], Available: https://aidsinfo.unaids.org/ [2020, January 19].

Ungerer, M., Bowmaker-Falconer, A., Oosthuizen, C., Phehane, V. & Strever, A. 2018. *The future of the Western Cape agricultural sector in the context of the Fourth Industrial Revolution Synthesis report*.

Vink, N. & Tregurtha, N. 2005. *Western Cape agricultural sector: structure , performance and future prospects*. [Online], Available: https://www.westerncape.gov.za/other/2005/10/final_first_paper_overview_agriculture.pdf.

Vyas, S. & Kumaranayake, L. 2006. Constructing socio-economic status indices: How to use principal components analysis. *Health Policy and Planning*. 21(6):459–468.

Walker, W. 2005. The strengths and weaknesses of research designs involving quantitative measures. *Journal of Research in Nursing*. 10(5):583–583.

Walther, B., Hossin, S., Townend, J., Abernethy, N., Parker, D. & Jeffries, D. 2011. Comparison of electronic data capture (EDC) with the standard data capture method for clinical trial data. *PLoS ONE*. 6(9).

Western Cape Governement. 2015. *Provincial Human Settlement Demand Profile*.

Van Wijk, E.C., Overberg, R.I., Kunst, A.E. & Harting, J. 2020. Opportunities for tailored support to implement smoke-free homes: A qualitative study among lower socioeconomic status parents. *International Journal of Environmental Research and Public Health*. 17(1).

Wolff, E.N. 2001. Wealth Distribution. In New York University, New York, USA *International Encyclopedia of the Social & Behavioral Sciences*. 16394–16401.

Yakubu, Y.A.M. & Schutte, D.W. 2019. Socio-economic determinants and environmental hygiene factors of female caregiver burden in two selected low-income communities in Cape Town, South Africa. *Pan African Medical Journal*. 34:1–9.

Appendix A: QASES observation tool

Housing

Scale	Description
0	 A poorly built shack with limited access to sanitation and possibly overcrowded. The shack could potential be small, with improper windows. Provides little protection from the elements.
1	 A well construct shack which tends to be on the larger side. It provides adequate ventilation and protection against the elements. Or a very small formal structure with informal structures added to it.

2	 Subsidised housing apartments and possibly well maintained hostels. These structures contain ablution facilities within the structure (may be shared facilities).
3	  Houses range from Reconstruction and Development Programme (RDP) housing to deteriorated privately owned housing.

4	 Well-maintained and fully secured privately owned houses

Assets

Scale	Description
0	Households with limited assets, e.g. only a bed/mattress, crockery, utensils, pots/pans, and a two burner stove.
1	Households with a few assets like a two burner stove or mini-oven with hobs, utensils, crockery, pots/pans, fridge, kettle, cell phone, a tubed television, a radio, beds and mattresses, etc.
2	Households with a moderate amount of assets like an electric stove, a microwave, a fridge, a kettle, utensils, crockery, pots/pans, washing machine, iron, a television, cell phone, and a tablet, beds, lounge suite of poor quality, etc.
3	Households that have most assets like an electric stove, a microwave, a fridge, a kettle, utensils, crockery, pots/pans, washing machine, iron, a flat screen/tubed television, one or more cell phone, and a tablet and a desktop computer, dining set, good quality lounge suite, beds, possibly a privately owned vehicle etc.
4	Households that have a great amount of assets e.g. one or more privately owned vehicles, utensils, pots/pans, crockery, desktop computers/ a laptop, a tablet, cell phones, one or more TVs, sound systems, electric stove, two-burner plates, mini-ovens with hobs, washing machine, microwave, tumble dryer, fridge, freezer, kettle, iron, any other gadgets (slow cooker, pressure cookers, flat irons and hairdryers, etc.), furniture, e.g. lounge suites, dining sets, beds/mattress, cupboards (kitchen and bedroom), tables/countertops, branded clothing, jewellery and toiletries available, etc.

Community Outlook

Scale	Description
0	A hopeless place of abject misery and brutality where people with no other options exist and have no possibility of escape.
1	Smelly, dirty, and generally objectionable where the humour is dark, people drink sullenly and anger saturates the air where community members who have fallen on hard times are slowly clawing their way out of – or are on the dissipating slide to defeat.
2	An established home to the oppressed who grind from one menially demeaning day to the next waiting for the meagre relief of weekends and the all too expensive festive/holiday season.
3	Residents cling to the tenuous hope that oppression and poverty are not inevitable (perhaps for them, but maybe not their children), with grand dreams of securing their place in the community and mean determination. A place of dashed dreams and firm resolve.
4	Although still part of the South African context of crime, vigilance, and inter-group tensions, this is a place where residents have a good chance of upward social mobility, can contentedly engage in pleasurable pastimes/hobbies (like travelling to the city for shopping at weekends), other community members might aspire to live here and residents look down their noses at other areas in the community.

Appendix B: PC0 Economic Activity and Food Security

DATASET LABEL	ECASELINK NAME	EXPORT RULE	SAS NAME	SAS LABEL	SAS TYPE	SAS TYPE SIZE	SAS FORMAT
Economic Activity and Food Security	ECWATERNR	Do Not Export					
Economic Activity and Food Security	ECITEM12	Do Not Export					
Economic Activity and Food Security	EX_EXPORT	Do Not Export					
Economic Activity and Food Security	RECORD_ID	Do Not Export					
Economic Activity and Food Security	PROTOCOL_ID	Do Not Export					
Economic Activity and Food Security	DOC_ID	Do Not Export					
Economic Activity and Food Security	SITE_ID	Do Not Export					
Economic Activity and Food Security	PATIENT_INITIALS	Do Not Export					
Economic Activity and Food Security	PAGE_NUMBER	Do Not Export					
Economic Activity and Food Security	VISITNUM	Do Not Export					
Economic Activity and Food Security	ROW_NO	Do Not Export					
Economic Activity and Food Security		Assign value from HH.HID	HID	Household	NUMERIC	8	
Economic Activity and Food Security	SITE_NO		SITE_NO	Site Number	CHARACTER	50	
Economic Activity and Food Security	SCREEN_NO		SCREEN_NO	Screening Number	CHARACTER	50	
Economic Activity and Food Security	VISIT		VISIT	Visit	NUMERIC	8	
Economic Activity and Food Security	DEVICE_ID		DEVICE_ID	Device Id	NUMERIC	8	

Economic Activity and Food Security	ENTRY_ID		ENTRY_ID	Entry Id	CHARACTER	50	
Economic Activity and Food Security	ENTRY_DATETIME		ENTRY_DATETIME	Entry Datetime	NUMERIC	8	DD/MM/YYYY
Economic Activity and Food Security		Assign value from TBL_PATIENT_ENTRY.PTID	PTID	Participant	NUMERIC	8	-
Economic Activity and Food Security	ECBUILD	1. If value='No Answer' or null assign value as '.M' Transformational Rules Single unit/brick structure on its own stand=1 Cluster/multi-unit=2 Traditional hut/structure made from traditional material=3 Flat in block of flats=4 Servant quarters=5 Caravan/tent=6 Worker's hostel=7 Shack=8 Other=9	ECBUILD	Which of the following best describes the main type of building that this household occupies?	NUMERIC	8	ECBUILD 1=Single unit/brick structure on its own stand 2=Cluster/multi-unit 3=Traditional hut/structure made from traditional material 4=Flat in block of flats 5=Servant quarters 6=Caravan/tent 7=Worker's hostel 8=Shack 9=Other
Economic Activity and Food Security	ECFLOOR	1. If value='No Answer' or null assign value as '.M' Transformational Rules Dirt/earth=1 Wood/plank=2 Parquet=3 Lino=4 Cement=5 Tile=6 Other=7	ECFLOOR	What is the main type of flooring for this household?	NUMERIC	8	ECFLOOR 1=Dirt/earth 2=Wood/plank 3=Parquet 4=Lino 5=Cement 6=Tile 7=Other

Economic Activity and Food Security	ECROOM	1. If value='No Answer' or null assign value as '.M'	ECROOM	How many living and sleeping rooms are there in this housing unit?	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECROOMU	1. If value='No Answer' or null assign value as '.M' 2. Assign value as null if ECROOM is null	ECROOMU	Rooms	CHARACTER	5	
Economic Activity and Food Security	ECROOMNR	1. If value='No Answer' or null assign value as '.M'	ECROOMNR	No Answer	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECWATER	1. If value='No Answer' or null assign value as '.M' Transformational Rules Piped indoors=1 Stand pipe/tap within plot=2 Communal tap=3 Borehole=4 Protected well=5 Unprotected/shallow well =6 river/dam/lake/pan=7 Bowser/tanker=8 Other=9	ECWATER	What is the main source of drinking water for this household?	NUMERIC	8	ECWATER 1=Piped indoors 2=Stand pipe/tap within plot 3=Communal tap 4=Borehole 5=Protected well 6=Unprotected/shallow well 7=river/dam/lake/pan 8=Bowser/tanker 9=Other
Economic Activity and Food Security	ECENERGY	1. If value='No Answer' or null' assign value as '.M' Transformational Rules No cooking done in household=1 Electricity (mains)=2 Electricity (individual solar)=3 Gas=4 Paraffin=5 Charcoal =6 Wood=7 Other=8	ECENERGY	What is the main source of energy used for cooking?	NUMERIC	8	ECENERGY 1=No cooking done in household 2=Electricity (mains) 3=Electricity (individual solar) 4=Gas 5=Paraffin 6=Charcoal 7=Wood 8=Other

Economic Activity and Food Security	ECITEM0	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM0	None	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM1	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM1	Working Cellphone	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM2	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM2	Bicycle	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM3	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM3	Motorcycle or scooter	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM4	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM4	Car/bakkie	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM5	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM5	Electricity to house	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM6	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM6	TV	NUMERIC	8	CHECK

Economic Activity and Food Security	ECITEM7	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM7	Fridge/freezer	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM8	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM8	Radio	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM9	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM9	Computer/Laptop	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM10	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM10	CD or MP3 player	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEM11	1. If value is null or ECITEMNR='No Answer' assign value as '.S' 2. If value is not null and ECITEMNR<>'No Answer' assign value as '1'	ECITEM11	Stereo/cassette /other music player	NUMERIC	8	CHECK
Economic Activity and Food Security	ECITEMNR	If value is not null assign value as '99'	ECITEMNR	No Answer	NUMERIC	8	NA

Economic Activity and Food Security	ECTOILET	1. If value='No Answer' or null assign value as '.M' Transfomation Rules Own flush toilet=1 Shared flush toilet=2 Own pit latrine=3 Shared pit latrine=4 Own VIP latrine=5 Shared VIP latrine=6 Pail/bucket=7 Communal chemical latrine=8 Bush=9 Other=10	ECTOILET	What is the main toilet facility used in your household?	NUMERIC	8	ECTOILET 1=Own flush toilet 2=Shared flush toilet 3=Own pit latrine 4=Shared pit latrine 5=Own VIP latrine 6=Shared VIP latrine 7=Pail/bucket 8=Communal chemical latrine 9=Bush 10=Other
Economic Activity and Food Security	ECINCOME	1. If value='No Answer' or null assign value as '.M' Transfomation Rules Regular employment (salary)=1 Farming=2 Trading=3 Fishing=4 Commercial sex work=5 Transportation=6 Artisan=7 Government grant=8 Other=9	ECINCOME	What is the main source of income for this household?	NUMERIC	8	ECINCOME 1=Regular employment (salary) 2=Farming 3=Trading 4=Fishing 5=Commercial sex work 6=Transportation 7=Artisan 8=Government grant 9=Other
Economic Activity and Food Security	ECGRANT	1. If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1	ECGRANT	Do you or anyone in your household receive any government grants?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3

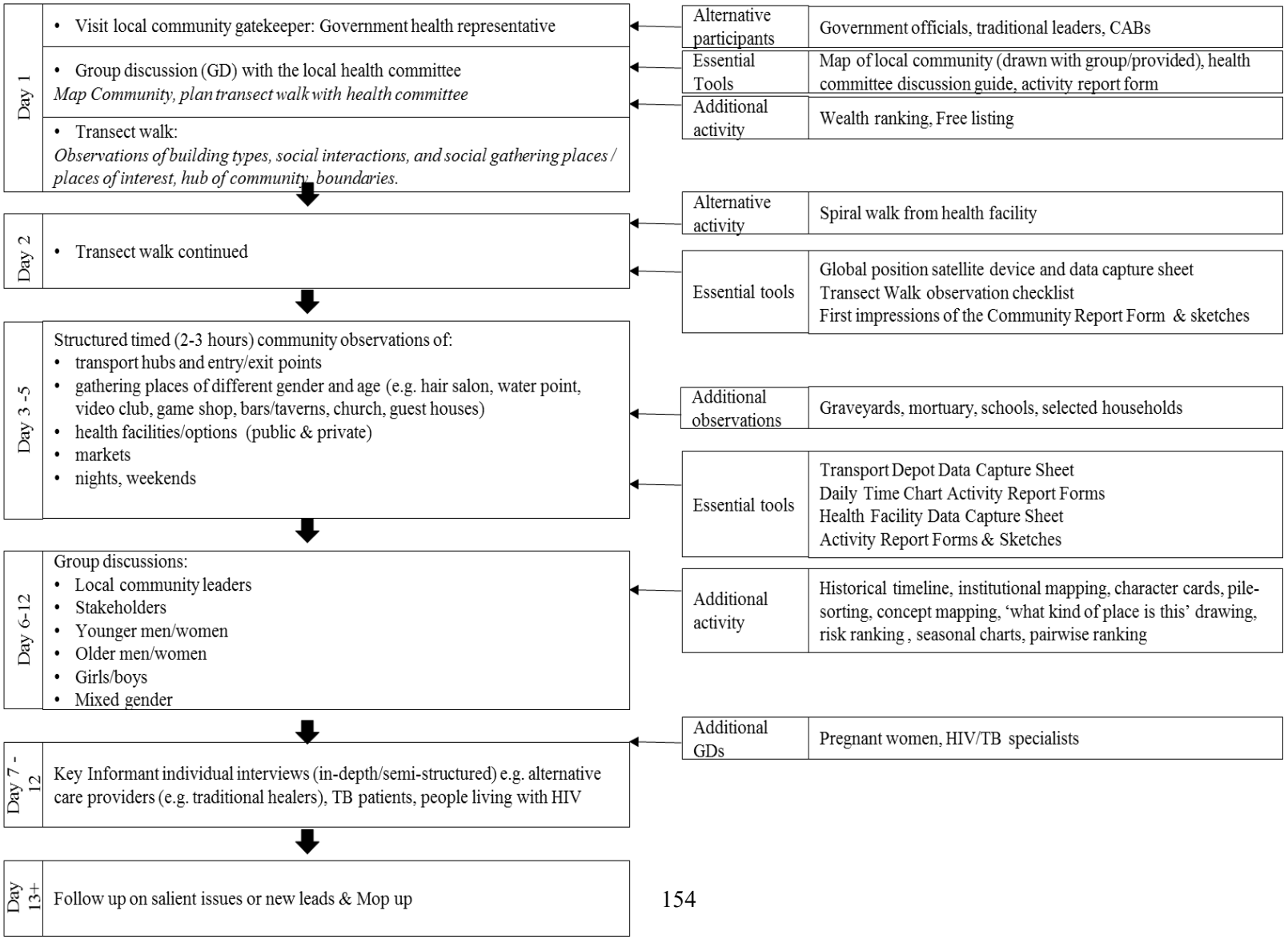
Economic Activity and Food Security	ECCHILD	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S' 3. if ECNR is null and value is not null assign value as '1'	ECCHILD	Child support	NUMERIC	8	CHECK
Economic Activity and Food Security	ECCHILDNUM	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S'	ECCHILDNUM	How many of each type of grant do you receive?	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECPENS	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S' 3. if ECNR is null and value is not null assign value as '1'	ECPENS	Old age pension	NUMERIC	8	CHECK
Economic Activity and Food Security	ECPENSNUM	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S'	ECPENSNUM	How many of each type of grant do you receive?	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECFOST	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S' 3. if ECNR is null and value is not null assign value as '1'	ECFOST	Foster care	NUMERIC	8	CHECK
Economic Activity and Food Security	ECFOSTNUM	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S'	ECFOSTNUM	How many of each type of grant do you receive?	NUMERIC	8	MISSKIP

Economic Activity and Food Security	ECDIS	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S' 3. if ECNR is null and value is not null assign value as '1'	ECDIS	Disability (disability dependency)	NUMERIC	8	CHECK
Economic Activity and Food Security	ECDISNUM	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S'	ECDISNUM	How many of each type of grant do you receive?	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECOTH	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S' 3. if ECNR is null and value is not null assign value as '1'	ECOTH	Other	NUMERIC	8	CHECK
Economic Activity and Food Security	ECOTHNUM	1. If ECGRANT='Yes' and ECNR<>'No Answer' and value is null assign value as '.M' 2. If ECGRANT<>'Yes' assign value as '.S'	ECOTHNUM	How many of each type of grant do you receive?	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECNR	1. If value is not null assign value as '99' 2. If ECGRANT<>'Yes' assign value as '.S'	ECNR	No Answer	NUMERIC	8	MISSKIP
Economic Activity and Food Security	ECFSMON	If ECFSMONNR<>'No Answer' and value is null assign value as '.M'	ECFSMON	During the last 12 months, how many months were there when your household did not have enough food to eat?	NUMERIC	8	MISSKIP

Economic Activity and Food Security	ECFSMONU	1. If ECFSMONNR<>'No Answer' and value is null assign value as '.M' 2. Assign value as null if ECFSMON is null	ECFSMONU	Months	CHARACTER	6	
Economic Activity and Food Security	ECFSMONNR	If value is not null assign value as '99'	ECFSMONNR	No Answer	NUMERIC	8	NA
Economic Activity and Food Security	ECRELIEF	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECRELIEF	In the last 12 months, did your household receive relief food or free food from government or other groups?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECREDUCE	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECREDUCE	In the last 12 months, did your household have to reduce the number of meals or food intake, because you did not have enough food?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECSHORT	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECSHORT	In the last 12 months, did your household have to eat food you would not ordinarily eat for meals, because of food shortage or lack of money?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECHHITEM	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECHHITEM	In the last 12 months, did your household have to reduce spending on other household items (e.g. soap, tissues), because of lack of money?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3

Economic Activity and Food Security	ECBORROW	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECBORROW	In the last 12 months, did your household have to borrow cash (kaloba, borrowing from friends) because of lack of money?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECBELONG	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECBELONG	In the last 12 months, did your household have to sell belongings, because of lack of money?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECPETTY	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECPETTY	In the last 12 months, did your household have to rely on petty vending/Piece work?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3
Economic Activity and Food Security	ECAWAY	If value='No Answer' or null assign value as '.M' Transformation Rules No=0 Yes=1 Don't Know=2	ECAWAY	In the last 12 months, did your household have to send household members away?	NUMERIC	8	YESNO Yes = 1 No= 2 Not Sure = 3 Don't Know = 3

Appendix C: BBS methods in sequence



Appendix D: Population Cohort informed consent

INFORMED CONSENT FORM – POPULATION COHORT
PARTICIPANT INFORMATION AND CONSENT FORM

Title of Research Study:	**Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa**
Protocol #:	HPTN 071, Version 1.0, 26 October 2012 DAIDS ID: 11865
Sponsor:	National Institute of Allergy and Infectious Diseases National Institute of Mental Health (U.S. National Institutes of Health) Office of the United States Global AIDS Coordinator Bill and Melinda Gates Foundation
Investigator of Record:	Nulda Beyers\Peter Bock

Research Site Address (es):

Site: Delft South Clinic **Address:** Cr Main Rd & Boyce St	**Site:** Bloekombos Clinic **Address:** Sam Nokasela Avenue	**Site:** Ikwhezi clinic **Address:** Simon Street Nomzame
Site: Town 2 Clinic (outreach) **Address:** c/o Zibonele and Manyano Street	**Site:** Kuyasa Clinic **Address:** Ntlazana Street, Khayelitsha	**Site:** Luvuyo Clinic **Address:** Hlela Road, Makaza
Site: Dalevale Clinic (outreach) **Address:** Symphony Avenue,	**Site:** Cloetesville Clinic **Address:** Tennant Street	**Site:** Wellington Clinic (outreach) **Address:** Wellington Municipality

Daytime telephone number(s): **021 983 9114**

24-hour contact number(s): **083 572 1470**

Participant Information and Consent Form
Please ask the study investigator or the study staff to explain any words or procedures that you do not clearly understand.

The purpose of this form is to give you information about the research study you are being asked to join. If you sign this form, you will be giving your permission to take part in the study. The form describes the purpose, procedures, benefits, and risks of the research study. You should take part in the study only if you want to do so. You may choose not to join the research project or withdraw from this study at any time. Choosing not to take part in this research will not in any way affect the health care or benefits that you or your family will receive. Please read this t Information and Consent Form and ask as many questions as needed. You should not sign this form if you have any questions that have not been answered to your satisfaction.

This study is being funded by the U.S. National Institutes of Health, the Office of the United States Global AIDS Coordinator, and the Bill and Melinda Gates Foundation

Your participation is voluntary
You do not have to take part in this study. If you decide today to take part in this research project, you may refuse to take part in any portion of the study or stop at any time without reducing or affecting any care that you receive at the health centers in your community.

Purpose of the Research in the Communities
The HPTN 071 or PopART study is testing a program to try to reduce HIV infection in a community like yours. Twenty one communities that include about 600,000 adults are included in this research (about 400,000 adults in twelve Zambian communities and 200,000 adults in nine South African communities, all located in the Western Cape).

In some communities, the level of care that people are used to will stay the same, in terms of HIV testing, and care of those who have HIV.

In other communities, to make HIV testing easier, community health care workers will go to all homes and will offer to do an HIV test on those wishing to have a test. (In South Africa children over the age of twelve can choose to have an HIV test without getting permission from their parents or guardians although it is better to first get consent from parents or guardians). For anyone infected with HIV, they will be offered to start taking drugs to treat HIV according to the standard treatment guidelines that are in place in the Western Cape. The health workers will visit every home again once a year for up to three more years to repeat the HIV testing and to refer people to care.

In other communities, health care workers will go to all houses offering HIV testing, as was just described. In these communities if someone over the age of 18 tests HIV positive however, they will be offered to start taking medicines to treat HIV right away. The health workers will visit every home again once a year for up to three more years to repeat the HIV testing and to refer people to care. Children under the age of 18 who test HIV positive will be offered care according to the standard treatment guidelines used in the Western Cape.

At the end of the study, the researchers will see if offering HIV tests in each household and offering people the chance to start HIV treatment right away has reduced the number of people with HIV infection in the community. They will also see if starting ART early has any negative effects on people's health.

Your community is one of the communities participating in this research. If health care workers are visiting homes in your community, you will notice that they provide some other information and services to people, but the most important thing is the testing and HIV treatment they offer.

In each community, around 2,700 people will be asked to participate in additional activities such as completing questionnaires and providing additional samples for laboratory testing. These questionnaires and tests will let the researchers understand how the community feels about the program and if the program is working. You have been selected to be one of the people from your community who we are asking to participate in these additional activities. That is why you are being asked to read this document.

What will happen during this study?
If you participate in this study, you will have up to four study visits: today, in 12 months, in 24 months, and possibly a final visit in 36 months. We will contact you to remind you about your visits. For example, we may call you or send a short text message (SMS). Today's visit will take approximately 2 hours. Future visits may be slightly shorter. Today we will:

- Ask you questions about a number of topics including you and your sexual practices, HIV testing, male circumcision, and how you and others feel about HIV.

- Some people may have to answer more sections than others. These people will be selected in a process similar to flipping a coin. The reason for this is that there is not enough time for everyone to answer all the questions.
    - Collect up to 15 mL blood (about 3 teaspoons) for HIV testing and other HIV-related tests as well as herpes simplex-2 testing. The results of these tests are for research purposes only and will not be given back to you. However if you agree to participate in the study, we will offer to perform an on-the-spot HIV test (in the form of a rapid test) at each visit, and will provide counseling if you would like to know the result of your test. If these tests say that you are positive for HIV, we will refer you for care at the local health center
    - The staff at the health center keep records of all their patients as part of their normal procedures. We would like to look at these medical records for any study participant who is HIV infected. Doing so will help us better understand how the study activities in the community are affecting the health of people diagnosed with HIV. If you agree to participate in this study, we will ask you for your permission to look at your records at the health center. This may include information collected by the community health workers if they are visiting homes in your community.

What are the possible risks or discomforts?
You may become embarrassed, worried or anxious when learning your HIV status and discussing sexual risk behavior and other topics. A trained staff member will help you deal with any feelings or questions you have. You may feel that being part of this study could lead to you feeling stigmatized or separated from our community

It is very unusual to have any problems from having a blood test but you may feel discomfort, dizzy, or even faint when your blood is drawn. Redness, pain, swelling, bruising may occur where the needle goes into your arm but this is rare.

What are the potential benefits?
During the study, you can decide if you would like to learn your HIV status and be provided with information on where to receive treatment and care services if needed. You will also be able to ask questions about your health.

In addition, knowledge gained from this study may help reduce the spread of HIV in the future and promote better health for you and your family as well as helping with acknowledgement and acceptance of HIV as a community-wide health problem.

Are there any alternatives to participation?
If you decide not to participate in this study, we will refer you to other places where you can receive an HIV test. If it is offered in your community, you can also receive testing from a health worker visiting your home during the study period.
How will my confidentiality and privacy be protected?
We cannot guarantee absolute confidentiality. However, we will do everything possible to protect your confidentiality if you join this study. We do this by giving you a study number and any information will be labeled with this number only, so people working in the health centers and laboratories will only see a number not your name, only the research staff will be able to link this number to your name. Your personal information (name, address, phone number) will be protected by the research staff. This information will not be used in any publication of information about this study.

To protect your privacy, you will meet with the researcher in a private area where others cannot overhear conversations with you.

People who may review your records include: Stellenbosch University Health Research Ethics Committee, local regulatory agencies, US National Institutes of Health (NIH), study staff, and study

monitors. Institutional Review Boards (IRBs) or Ethics Committees (ECs) are committees that watch over the safety and rights of research participants.

Storage of blood samples for future use
In addition to the testing that is part of the PopART study, the PopART team is asking for permission to store any leftover specimens for other future research related to HIV and other diseases. Through these studies, researchers hope to find new ways to detect, treat, and maybe prevent or cure these health problems. There is currently no cure for HIV.

- We will **not** use your stored blood for **human genetic research** (to study your genes). If you are HIV positive, we may use the blood to study the HIV virus.
- Some of your blood may be stored at the Desmond Tutu TB Centre in South Africa. Some of your blood will be shipped and securely stored outside of the country in the USA.
- Your privacy is very important to us and we will make every effort to protect it.
    - We will remove your name and other identifiers from your sample and information, and replace them with a code number. We will keep the list that links the code number to your name separate from your sample and information and this list will only be accessible to the Principal Investigator and senior research staff at Desmond Tutu TB Center. They will sign an agreement to keep your identity a secret.
    - Researchers who study your samples and information will not know who you are. They must also agree that they will not try to find out who you are.

You can refuse use and/or storage of your leftover samples and still take part in this study. You will be able to mark your decision at the end of this form.

What happens if I am injured by participating in this study?
It is very unlikely that you could be injured as a result of participating in this study. However, if you are injured while participating in this study, you will be given immediate treatment for your injuries. You will not have to pay for this care. There is no program for compensation through the United States NIH, however Stellenbosch University does have an insurance cover for compensation of serious research related injury. You will not be giving up any of your legal rights by signing this Participant Information and Consent Form.

What are some reasons why I may be withdrawn from this activity without my consent?
You may be withdrawn from the study without your consent for the following reasons:

- The research study, or this part of the study, is stopped or canceled
- The study staff feels that completing the study or this part of the study would be harmful to you or others

This study has been approved by the **Health Research Ethics Committee at Stellenbosch University** and will be conducted according to the ethical guidelines and principles of the international Declaration of Helsinki, South African Guidelines for Good Clinical Practice and the Medical Research Council (MRC) Ethical Guidelines for Research.

Persons to Contact for Problems or Questions
If you have any questions about your participation in this research study, your rights as a research subject, or if you feel that you have experienced a research-related injury, contact:

1. Dr Peter Bock, Co-Principal Investigator, Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, University of Stellenbosch. Telephone: 021 9389062. Email: peterb@sun.ac.za

2. Principal Investigator: Nulda Beyers, Desmond Tutu TB Centre, Department of Paediatrics and Child Health, Faculty of Medicine and Health Sciences, University of Stellenbosch. Telephone: 021 938 9114. Email:nb@sun.ac.za

3. Mr Franklin Weber, HREC coordinator, Health Research Committee 1, Stellenbosch University Health Research Ethics Committee, Tygerberg Campus. Telephone: 021 938 9657.

PARTICPANTS STATEMENT OF CONSENT

Population Effects of Antiretroviral Therapy to Reduce HIV Transmission (PopART): A cluster-randomized trial of the impact of a combination prevention package on population-level HIV incidence in Zambia and South Africa

- I have been given sufficient time to consider whether to take part in this study.
- My taking part in this research study is voluntary. I may decide not to take part or to withdraw from the research study at any time without penalty or loss of benefits or treatment to which I am entitled.
- The research study may be stopped at any time without my consent.
- I have had an opportunity to ask my study investigator questions about this research study. My questions so far have been answered to my satisfaction.
- I have been told how long I may be in the research study.
- I have been informed of the procedures and tests that may be performed during the research study.
- I have been told what the possible risks and benefits are from taking part in this research study. I may not benefit if I take part in this research study.
- I do not give up my legal rights by signing this form.
- I have been told that before any study related procedures being performed, I will be asked to voluntarily sign this Participant Information and Consent Form.
- I will receive a signed and dated copy of this Information and Consent Form.

If you have either read or have heard the information in this Participant Information and Consent Form, if all of your questions have been answered, and if you agree to take part in the study, please print and sign your name and write the date on the line below.

Insert barcode label here

Fingerprinting

We would like to use an electronic device to scan your fingerprint at all visits. Your fingerprint will be used to identify you correctly at follow up visits. Your fingerprint will not be used for any other purpose. You may choose not to have your fingerprint taken, in which case you will be asked a few additional questions to confirm your identify at follow up visits.

_____ My initials indicate that I agree to having my fingerprint used for this study.

Access of Data from Health Center

_____ My initials indicate that I agree to allow my records at the health center to be accessed and used for this study.

_____ I do not agree to allow my health care records to be accessed and used for this study.

Storage of leftover specimens

_____ My initials indicate that I agree to allow my leftover samples to be used for future research.

_____ I do not agree to allow my leftover samples to be used for future research.

I voluntarily agree to take part in this research study.

----------	----------	----------
Subject's Name and Surname	**Subject's Signature**	**Date (dd/mm/yyyy) (print)**

I certify that the information provided was given in a language that was understandable to the subject.

----------	----------	----------
Name and surname of Study Staff Consent Discussion (print)	**Study Staff Signature**	**Date (dd/mm/yyyy)**

----------	----------	----------
Witness' Name and surname (As appropriate) (print)	**Witness' Signature**	**Date (dd/mm/yyyy)**

Appendix E: HPTN 071 (PopART) clearance letter for HREC

DESMOND TUTU TB CENTRE · SENTRUM · IZIKO
UNIVERSITY / STELLENBOSCH / UNIVERSITEIT

Every breath counts

22 February 2019

Dr Barsdorf
Health Research Ethics Committee
Division for Research Development
Stellenbosch University
Cape Town

Dear Dr Barsdorf

Re: Ms Melissa Nel's application for secondary analysis of HPTN 071 (PopART) [N12/11/074] data to be included in her MPhil in Transdisciplinary Health and Development Studies dissertation.

Many thanks for considering this letter. Ms Nel is a staff member of HPTN 071 (PopART) at the Desmond Tutu TB Centre. She has asked the trial PIs for permission to use data collected in the trial toward an MPhil degree – proposal entitled "The efficiency and accuracy of a rapid qualitative tool to ascribe socio-economic status in HPTN 071 (PopART) study communities in South Africa – a cross-sectional study". She will be supervised by two senior DTTC and HPTN 071 (PopART) staff; Graeme Hoddinott and Rory Dunbar. We have reviewed her proposal and are happy to say that it falls within the scope of HPTN 071 (PopART), will make a valuable independent contribution to science that is not otherwise planned, and that the trial leadership have agreed to support her access to the data for the aim/objectives specified.

We trust that you find this in order and are happy to answer any queries by HREC either via email or in person.

Sincerely

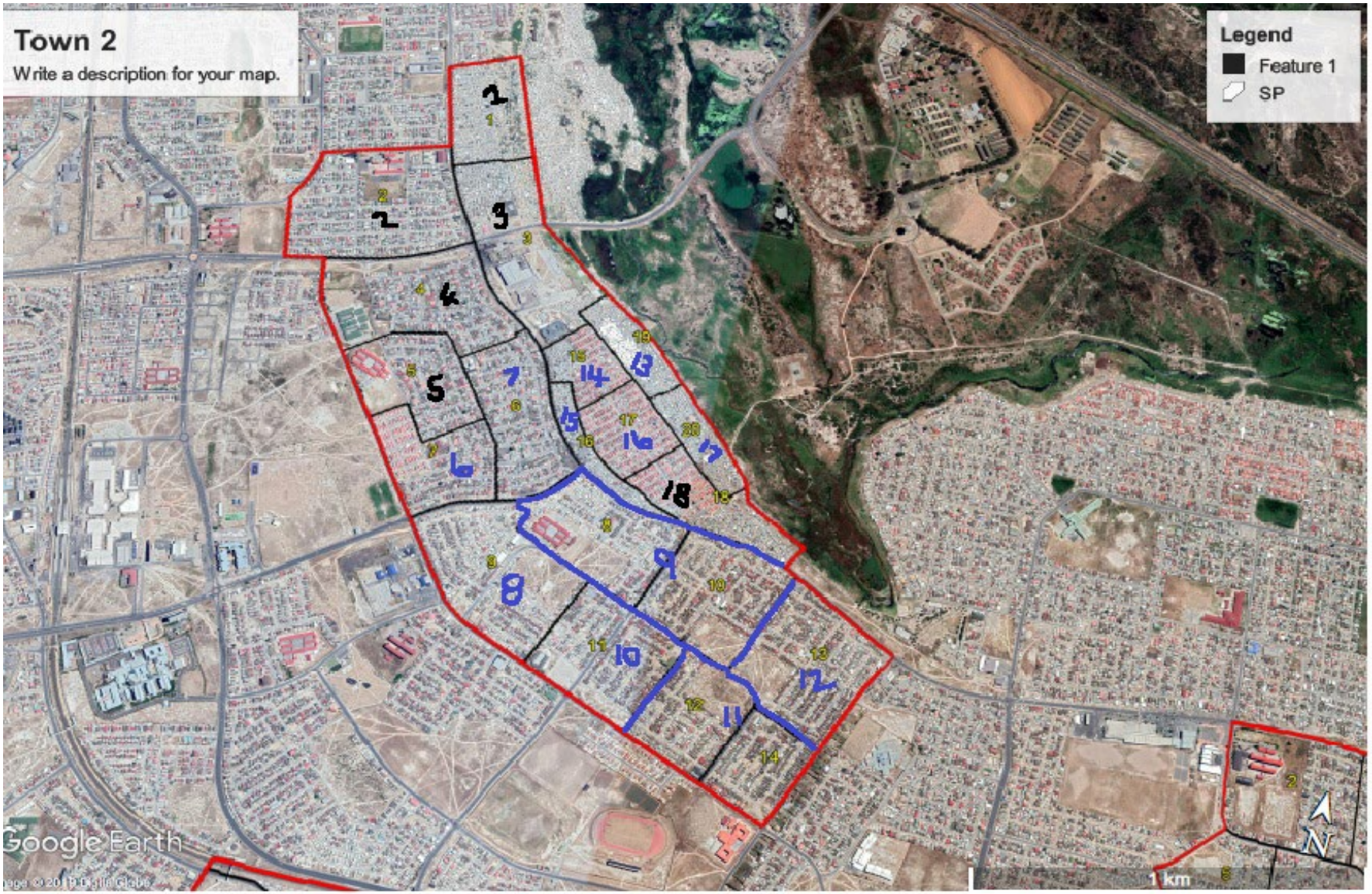
Nulda Beyers (PI)

Peter Bock (co-PI)

UNIVERSITEIT·STELLENBOSCH·UNIVERSITY
jou kennisvennoot • your knowledge partner

Desmond Tutu TB Sentrum ⁜ Centre ⁜ Iziko
Departement Pediatrie en Kindergesondheid ⁜ Department of Paediatrics and Child Health
Fakulteit Geneeskunde en Gesondheidswetenskappe ⁜ Faculty of Medicine and Health Sciences
241, Cape Town 8000 ☎ (27+21) 938 9812, Faks ⁜ Fax: (27+21) 938 9719, Suid Afrika ⁜ South Africa
Direkteur ⁜ Director: Prof Anneke Hesseling (annekeh@sun.ac.za)

Appendix F: Reworked map of aligning PC and BBS zones



Appendix G: Stem and leaf plot (identified outliers)

Descriptives

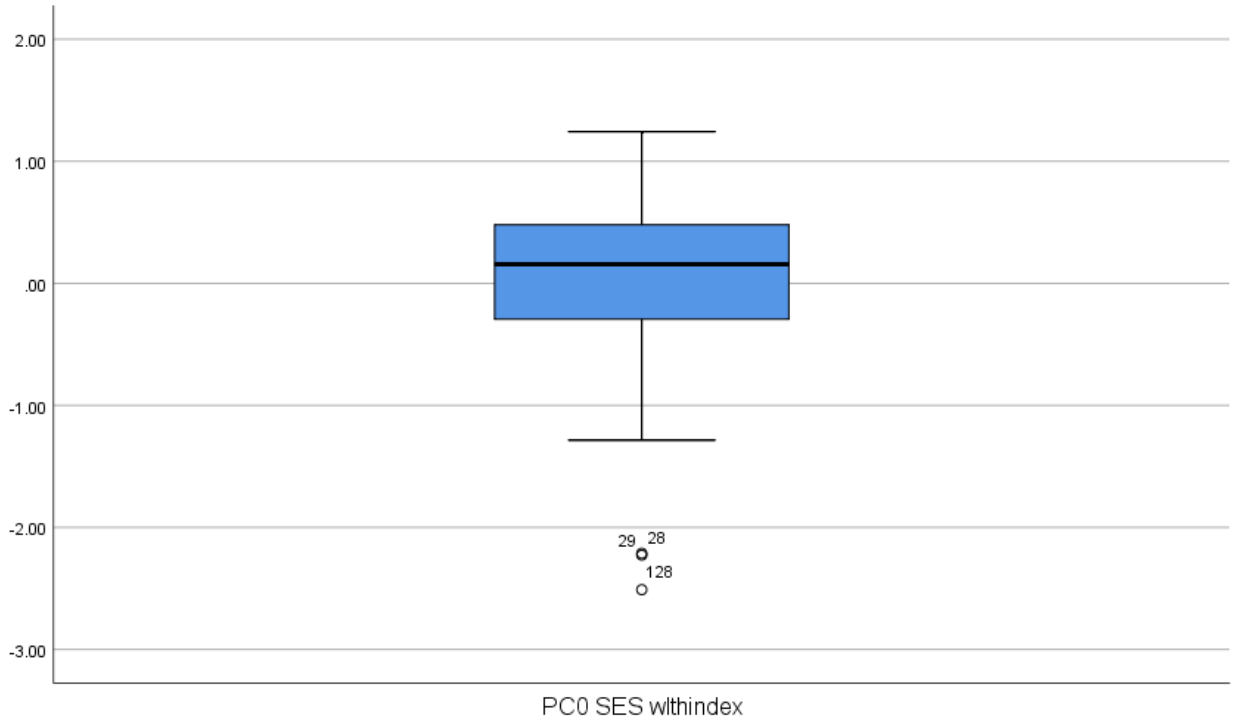
			Statistic	Std. Error
PC0 SES wlthindex	Mean		.067651290284223	.052424172251416
	95% Confidence Interval for Mean	Lower Bound	-.035969021434703	
		Upper Bound	.171271602003149	
	5% Trimmed Mean		.106153624556581	
	Median		.157103277060767	
	Variance		.399	
	Std. Deviation		.631270628380330	
	Minimum		-2.5101505901147100	
	Maximum		1.2430777474340200	
	Range		3.7532283375487294	
	Interquartile Range		.7900951927831250	
	Skewness		-1.200	.201
	Kurtosis		2.776	.400

Extreme Values

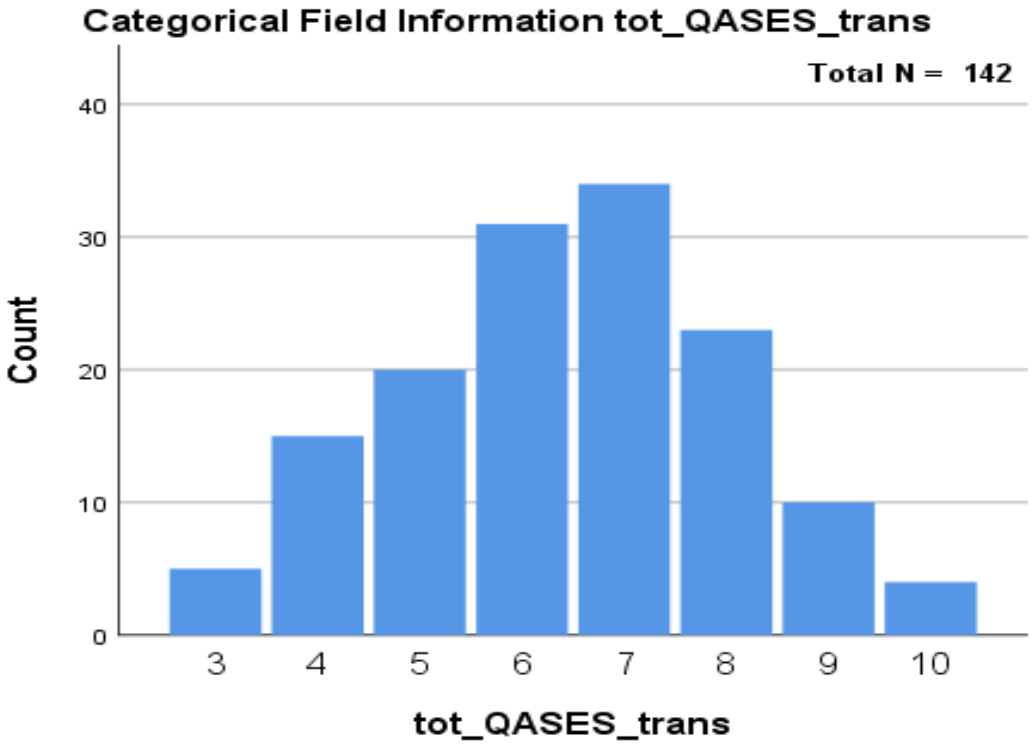
			Case Number	Value
PC0 SES wlthindex	Highest	1	99	1.2430777474340200
		2	97	1.1908925693871400
		3	121	1.0789866881961600
		4	139	1.0748920679713300
		5	98	1.0094924785437100
	Lowest	1	128	-2.5101505901147100
		2	29	-2.2270707332726700
		3	28	-2.2124867801730700
		4	30	-1.2851018225632300
		5	18	-1.1133926247842400

```
PC0 SES wlthindex Stem-and-Leaf Plot

 Frequency     Stem &  Leaf

     3,00 Extremes    (=<-2,2)
     1,00       -1 .  2
     3,00       -1 .  111
     5,00       -0 .  88889
     5,00       -0 .  66677
    12,00       -0 .  444444555555
    13,00       -0 .  2222223333333
    14,00       -0 .  00000000000111
    21,00        0 .  000000000111111111111
    26,00        0 .  22222222223333333333333333
    14,00        0 .  44444455555555
```

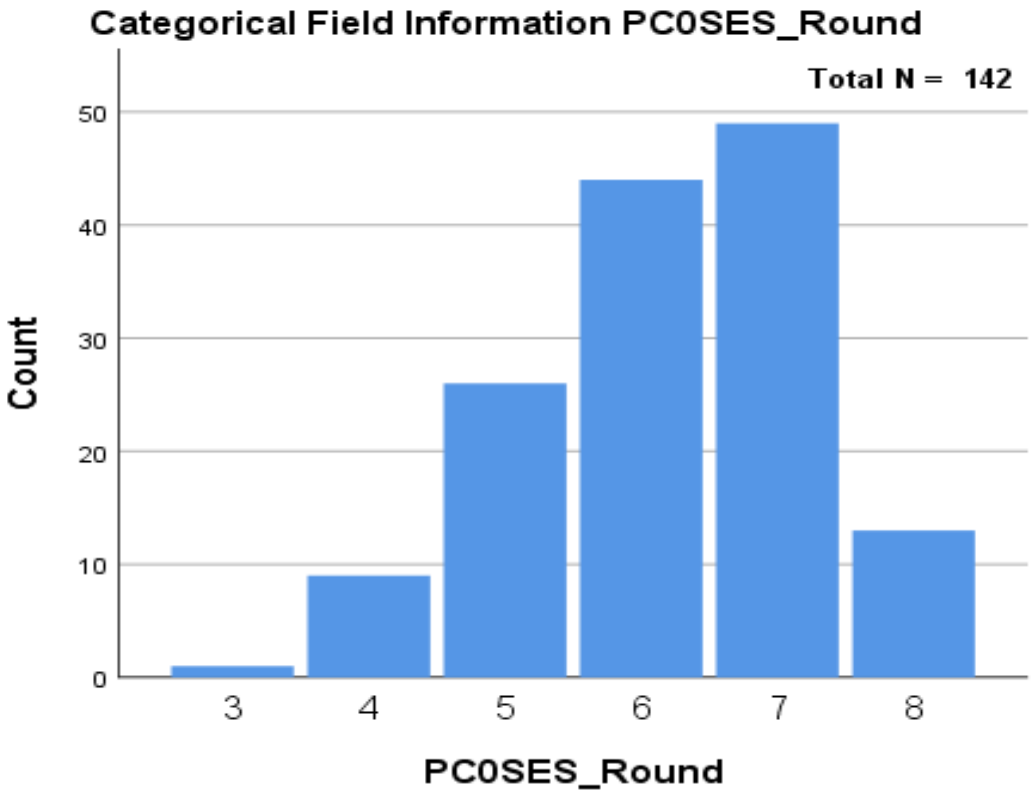
```
   18,00        0 .  666666666777777777
    4,00        0 .  8899
    5,00        1 .  00001
    1,00        1 .  2

Stem width:  1,000000
Each leaf:        1 case(s)
```



Appendix H: Transformed data of QASES and PC0



Outliers were first removed. Furthermore, the data scores were manually manipulated to fit better to the PC0 SES index. The central values (5, 6, 7, 8) were left untouched. To the left, +2 were added to values 1 and 2 and +1 were added to values 3 and 4. To the right, -2 were subtracted from values 11 and 12, and -1 were subtracted from values 10 and 11. This was done to increase the precision of the normal distribution.



PC0 data were changed from interval to ordinal. I applied the round formula on excel to the PC0 SES scores in order to make the values whole numbers. Then I added 3 to each score so that the values become positive numbers. The formula is – Round((PC0+3)*2).

Appendix I: Cross-tabulation of transformed total QASES and PC0 SES (matching of scores)

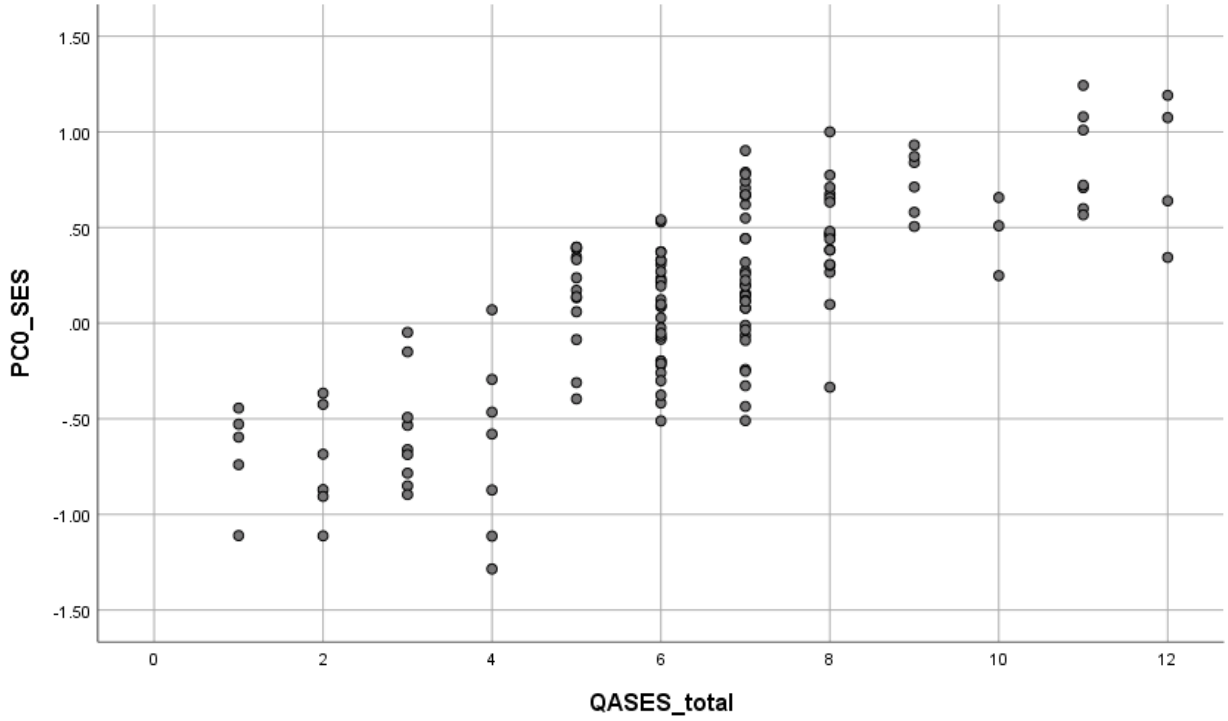
tot_QASES_trans * PC0SES_trans Crosstabulation

Count

		PC0SES_trans						
		3	4	5	6	7	8	Total
tot_QASES_trans	3	0	1	4	0	0	0	5
	4	0	6	7	2	0	0	15
	5	1	2	5	7	5	0	20
	6	0	0	5	18	8	0	31
	7	0	0	4	15	12	3	34
	8	0	0	1	1	16	5	23
	9	0	0	0	1	6	3	10
	10	0	0	0	0	2	2	4
Total		1	9	26	44	49	13	142

Appendix J: Scatter plot and correlation test of total QASES and PC0 (outliers removed)

The scatter plot indicates a strong, positive, linear relationship/association between X (total QASES) and Y (PC0 SES) with outliers excluded. The exclusion of the main outliers as calculated by the stem and leaf graph (see Appendix G).



Appendix K: Ethical Clearance Approval letter

Approval Notice

New Application

16/04/2019

Project ID :9068

HREC Reference #: S19/01/029

Title: The efficiency and accuracy of a rapid qualitative tool to ascribe socio-economic status in HPTN 071 (PopART) study communities in South Africa - a cross-sectional study

Dear Miss Melissa Nel,

The **Response to Modifications** received on 07/04/2019 14:59 was reviewed by members of **Health Research Ethics Committee 2 (HREC2)** via **expedited** review procedures on 16/04/2019 and was approved.

Please note the following information about your approved research protocol:

Protocol Approval Period: This project has approval for 12 months from the date of this letter.

Please remember to use your **Project ID [9068]** on any documents or correspondence with the HREC concerning your research protocol.

Please note that the HREC has the prerogative and authority to ask further questions, seek additional information, require further modifications, or monitor the conduct of your research and the consent process.

After Ethical Review

Translation of the informed consent document(s) to the language(s) applicable to your study participants should now be submitted to the HREC. Please note you can submit your progress report through the online ethics application process, available at: Links Application Form Direct

Link and the application should be submitted to the HREC before the year has expired. Please see *Forms and Instructions* on our HREC website (www.sun.ac.za/healthresearchethics) for guidance on how to submit a progress report.

The HREC will then consider the continuation of the project for a further year (if necessary). Annually a number of projects may be selected randomly for an external audit.

Provincial and City of Cape Town Approval

Please note that for research at a primary or secondary healthcare facility, permission must still be obtained from the relevant authorities (Western Cape Departement of Health and/or City Health) to conduct the research as stated in the protocol. Please consult the Western Cape Government website for access to the online Health Research Approval Process, see: https://www.westerncape.gov.za/general-publication/health-research- approval-process. Research that will be conducted at any tertiary academic institution requires approval from the relevant hospital manager. Ethics approval is required BEFORE approval can be obtained from these health authorities.

We wish you the best as you conduct your research.

For standard HREC forms and instructions, please visit: Forms and Instructions on our HREC website

https://applyethics.sun.ac.za/ProjectView/Index/9068

If you have any questions or need further assistance, please contact the HREC office at 021 938 9677.

Yours sincerely,

Mr. Francis Masiye,

HREC Coordinator

Health Research Ethics Committee 2 (HREC2).

National Health Research Ethics Council (NHREC) Registration Number: REC-130408-012 (HREC1)·REC-230208-010 (HREC2)

Federal Wide Assurance Number: 00001372
Office of Human Research Protections (OHRP) Institutional Review Board (IRB) Number: IRB0005240 (HREC1)·IRB0005239 (HREC2)

The Health Research Ethics Committee (HREC) complies with the SA National Health Act No. 61 of 2003 as it pertains to health research. The HREC abides by the ethical norms and principles for research, established by the World Medical Association (2013). Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects;the South African Department of Health (2006). Guidelines for Good Practice in the Conduct of Clinical Trials with Human Participants in South Africa (2nd edition); as well as the Department of Health (2015). Ethics in Health Research: Principles, Processes and Structures (2nd edition).

The Health Research Ethics Committee reviews research involving human subjects conducted or supported by the Department of Health and Human Services, or other federal departments or agencies that apply the Federal Policy for the Protection of Human Subjects to such research (United States Code of Federal Regulations Title 45 Part 46); and/or clinical investigations regulated by the Food and Drug Administration (FDA) of the Department of Health and Human Services.

Appendix L: Stellenbosch University base remuneration levels 2020

UNIVERSITEIT • IYUNIVESITHI • STELLENBOSCH • UNIVERSITY

100

BASE REMUNERATION LEVELS 2020

The new base remuneration levels (BRL) effective from 1 January 2020, are as follows:

Professional, administrative & support staff

Job Grade	Remuneration Level
Job Grade 19	75 600 – 84 000
Job Grade 15 - 17	108 900 – 121 000
Job Grade 14	130 500 - 145 000
Job Grade 13	156 600 – 174 000
Job Grade 12	188 100 – 209 000
Job Grade 11	244 800 – 272 000
Job Grade 10	317 700 – 353 000
Job Grade 9	382 500 - 425 000
Job Grade 8	475 200 - 528 000
Job Grade 7	586 800 – 652 000
Job Grade 6	757 800 – 842 000
Job Grade 5	1 012 262 – 1 124 736
Job Grade 4	1 312 222 – 1 458 024

Academic

Job Grade	Remuneration Level
P9: Junior lecturer	442 035 – 491 150
P8: Lecturer	614 827 – 683 142
P7: Senior lecturer	777 967 – 864 407
P6: Associate professor	941 501 – 1 046 112
P5: Professor	1 171 848 - 1 302 053
P4: Distinguished professor	1 312 222 - 1 458 024

Base remuneration levels (January 2020)

saam vorentoe • masiye phambili • forward together

Menslike Hulpbronne | Human Resources
Privaat Sak X1, Private Bag X1, Matieland, 7602, Suid-Afrika | South Africa
Tel: +27 21 808 2733 | Faks | Fax: +27 21 808 2464 | E-pos | Email: sun-e-hr@sun.ac.za